

UNIVERSITY OF CANTERBURY

Augmenting Immersive Cinematic Experience/Scene with User Body Visualization

Author:

Joshua Yang CHEN

Supervisor:

Dr. Christoph BARTNECK

Co-supervisor:

Dr. Gun LEE

Co-supervisor:

Prof. Mark BILLINGHURST

*A thesis submitted in fulfilment of the requirements
for the degree of Master of Human Interface Technology*

in the

Human Interface Technology Laboratory New Zealand
College of Engineering

February 2016

Declaration of Authorship

I, Joshua Yang CHEN, declare that this thesis titled, 'Augmenting Immersive Cinematic Experience/Scene with User Body Visualization' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

“As you get into new technology, you’ll know exactly what to do. And your work will not be about the technology. It will be about connecting and entertaining people.”

John Lasseter

UNIVERSITY OF CANTERBURY

Abstract

College of Engineering

Master of Human Interface Technology

Augmenting Immersive Cinematic Experience/Scene with User Body Visualization

by Joshua Yang CHEN

For the proposed masters thesis project, I conducted research into how computer graphics and emerging interactive technologies can be combined to create new immersive home entertainment experiences. It has the potential to change how people engage with the film from watching to experiencing it. Recent advances in hardware technology have led consumer level head-mounted displays to be ideal for immersive visualization providing wide-angle 3D stereo viewing. This project's involvement is to develop technology for blending the body of the user and the surrounding environment into a cinematic scene; doing so would provide the perception that the human viewer and digital content are in the same space. As a home entertainment experience, the system has to be simple, portable and easy to set up like a home video game console. The setup includes a head mounted display, an optical sensor to capture information about the user's environment, and some form of keying technology or technique to composite the user's body into a movie scene. Further investigation went into exploring different characteristics of 360 spherical panoramic movies and its effects with user embodiment on immersion; and also the type of control over blending the user body and surrounding environment that people would prefer; and lastly the type of interface that would be used to control this blending. Results showed a significant difference in the sense of presence and level of user engagement between when body visualization is present and when it is not present. It also showed that people preferred either shared or manual control with automated control the least. Implications of the results with valuable user feedback questioned how user embodiment should be presented and interaction should be supported now that the virtual environment is not any normal computer generated setting, but a cinematic virtual world driven by a story behind it. A set of guidelines is presented at the end with the target of a home entertainment system in mind.

Acknowledgements

The author wishes to express sincere appreciation to project supervisor Dr. Bartneck for his guidance and supervision. The author also would like to thank secondary supervisor Dr. Lee for being an inspiration and all his help with valuable advice, in both designing the application system and analysing the study data. Many thanks go out to Professor Billinghurst for his support and encouragement throughout this thesis. In addition, the author would also like to thank students and staff of the Human Interface Technology Laboratory New Zealand for their help in getting through the tough times and setbacks, and for creating a comfortable and enjoyable research environment. The author is extremely grateful to the HDI²4D joint research collaboration funded as part of the New Zealand - Korea Strategic Research Partnership, for sponsoring this research and the friendly international family. Lastly the author would like to thank his family and relatives back home for their love and sacrifice.

Contents

Declaration of Authorship	i
Abstract	iii
Acknowledgements	iv
Contents	v
List of Figures	viii
Abbreviations	x
1 Introduction	1
2 Background	6
2.1 Computer Graphics, Game Engines	6
2.2 Image Composition, Augmented Reality	7
2.3 Realism, Lighting, Augmented Virtuality	12
2.4 Head-mounted displays, Virtual Reality	14
2.5 Virtual environments, Sense of presence, Self-avatar	15
2.6 Critique	17
2.7 Conclusion	19
3 Design	21
3.1 Inspiration and Idea Generation	21
3.1.1 Brainstorm	21
3.1.2 User and Use case	24
3.1.3 Research Questions	25
4 Implementation and Prototype	28
4.1 The Hardware Requirements	28
4.1.1 Time-of-flight RGB-D camera	28
4.1.2 Head-mounted Display	30
4.1.3 Desktop PC	30
4.2 The Software	30
4.2.1 Unity 5	31
4.2.2 SoftKinetic iisu TM	31

4.2.3	Others	34
4.3	Prototype System Implementation	36
4.3.1	Extend UV map capabilities by building own depth-to-color mapping	38
4.3.2	Changing viewable depth through head-shaking and gestures	41
4.3.3	Attempts at parallel processing	43
4.3.4	Increasing the camera's field of view by extending the camera distance	44
4.3.5	UI for Implemented Gestures System	45
4.3.6	Viewable distance reading at the bottom	49
4.4	Summary	50
5	User Evaluation	52
5.1	Evaluation Goal	52
5.2	Focus Group	53
5.3	TEDx Conference	54
5.4	User Experiment Design	59
5.4.1	Hypothesis	61
5.4.2	Materials	62
5.4.3	Procedures	68
5.4.4	Measurements	69
5.4.5	Participants	71
6	Results and Discussion	73
6.1	Results	73
6.1.1	Quantitative Measures	73
6.1.1.1	IPQ	74
6.1.1.2	Engagement	76
6.1.1.3	Session 1 Post-experiment	77
6.1.1.4	Session 2 Post-experiment	86
6.1.2	Qualitative Measures	88
6.1.2.1	Session 1	89
6.1.2.2	Session 2	89
6.2	Discussion	90
6.2.1	Exclusion of participants	90
6.2.2	Further Discussion	95
6.3	Participants-feedback-inspired Future Improvements	97
6.4	Conclusions	99
7	Conclusion and Future Work	101
7.1	Conclusion	101
7.1.1	System with a focus on Home Entertainment	104
7.1.2	Future of 360 Film-making	105
7.2	Contribution	105
7.3	Future Work	108
7.3.1	Application with Gestures	108
7.3.2	Cinematic VE Experience - individual or group?	110
7.3.3	Extending FOV	111

7.3.4	Head Mounted Displays	111
7.3.5	Unreal Engine 4	113
7.3.6	Light Fields - The Future of VR-AR-MR -Light Fields	113
7.3.7	NextVR	114
7.3.8	Future User Studies	115
 A Appendix A: Information Sheet and Consent Form		117
 B Appendix B: Questionnaires		118
 Bibliography		119

List of Figures

1.1	Virtual Continuum	2
2.1	AR-Rift	8
2.2	Color cue hand tracking	9
2.3	Color chroma keying 1	10
2.4	Color chroma keying 2	11
2.5	Koola	13
2.6	Koola2	14
3.1	Concept Scenario Base	22
3.2	Concept Idea Flow Chart	23
4.1	Hardware set-up	29
4.2	iisu engine layers	33
4.3	iisu CI Layers	33
4.4	iisu CI Posing Gestures	35
4.5	iisu CI Moving Gestures	36
4.6	System screenshot	37
4.7	System Prototype Architecture	39
4.8	SoftKinetic DS325	40
4.9	Head-shaking flow chart	41
4.10	Extended Mount Prototype	45
4.11	Text-Based UI	46
4.12	Text-Based UI 2	47
4.13	Graphics-based UI ThumbsUp	47
4.14	Graphics-based UI OpenHand	48
4.15	Graphics-based UI ThumbsDown	48
4.16	Graphics-based UI Fading	49
4.17	Graphics-based UI Wave 1	50
4.18	Graphics-based UI Wave 2	51
5.1	TEDx Demo	54
5.2	TEDx Likert Summary Body Blending	56
5.3	TEDx Type of Movie Preference	57
5.4	TEDx 1st 3rd Person Viewpoint Preference	57
5.5	TEDx Moving Fixed Viewpoint Preference	58
5.6	TEDx Likert Responses Head Shaking	58
5.7	TEDx Visualization Control Preference	59
5.8	Movie with 1st Person Perspective	63

5.9	Movie with 3rd Person Perspective	63
5.10	2nd Movie with 3rd Person Perspective	64
5.11	3rd Movie with 3rd Person Perspective	64
5.12	Movie with a Virtual Body	65
5.13	Movie with No Virtual Body	66
5.14	Movie with Moving Perspective	66
5.15	Most Preferred Movie Genre	72
5.16	Least Preferred Movie Genre	72
6.1	Excluded Participants IPQ ratings	74
6.2	Participants IPQ ratings	75
6.3	Session 1 Conditions Legend	75
6.4	Participants IPQ Box-plot Likert Ratings	75
6.5	IPQ Statistics Results ART 2-way ANOVA	75
6.6	Excluded Participants Engagement ratings	76
6.7	Participants Engagement ratings	76
6.8	Session 1 Conditions Legend 2	77
6.9	Participants Engagement Box-plot Likert Ratings	77
6.10	Engagement Statistics Results ART 2-way ANOVA	77
6.11	Post-experiment 1 Qn2 Frequency Bar-chart	78
6.12	Post-experiment 1 Qn2 Statistics	78
6.13	Post-experiment 1 Qn4 Frequency Bar-chart	80
6.14	Post-experiment 1 Qn4 Statistics	80
6.15	Post-experiment 1 Qn6 Frequency Bar-chart	82
6.16	Post-experiment 1 Qn7 Frequency Bar-chart	82
6.17	Post-experiment 1 Qn9 Frequency Bar-chart	83
6.18	Post-experiment 1 Qn10 Frequency Bar-chart	83
6.19	Post-experiment 1 Qn11 Frequency Bar-chart	84
6.20	Post-experiment 1 Qn12 Frequency Bar-chart	84
6.21	Post-experiment 1 Qn13 Frequency Bar-chart	85
6.22	Post-experiment 1 Qn13 Chi-square Goodness of Fit Statistics	85
6.23	Post-experiment 2 Qn2 Frequency Bar-chart	86
6.24	Post-experiment 2 Qn4 Table	87
6.25	Purposeful Interactions Example 1	92
6.26	Purposeful Interactions Example 2	92
6.27	Purposeful Interactions Example 3	93
6.28	Purposeful Interactions Example 4	93
6.29	Purposeful Interactions Example 5	94
7.1	Clash of clans	109
7.2	Clash of clans 2	110
7.3	Light Fields Idea	114
7.4	Light Fields NextVR Idea	115

Abbreviations

AR	A ugmented R eality
AV	A ugmented V irtuality
VR	V irtual R eality
CG	C omputer G raphics
FOV	F ield O f V iew
FPS	F rame P er S econd
GPU	G raphics P rocessing U nit
HMD	H ead M ounted D isplay
DK2	D evelopment K it 2
RGB-D	R ed G reen B lue D epth
VE	V irtual E nvironment

Chapter 1

Introduction

This thesis explores how a simple setup of a head-mounted display (HMD) and a RGB-D camera powered on a desktop machine can be used for future home cinematic entertainment. In the approaching times, we believe that everyone would be able to experience the whole new world of technology and our world combined. Augmented Virtuality (AV) is the merging of the real world into the virtual world, and many research has been done using it. However most require laborious and expensive set-ups, non-existent to the common regular household. Our idea would explore the options of system portability, its performance, and understanding how would the virtual environments of 360 movie content be properly supported on the system.

The huge potentials of three dimensional Computer Graphics (CG), Virtual Reality (VR), Augmented Reality (AR), Interactive technologies, and Digital Film-making combined creates endless possibilities to how new immersive entertainment experiences would be in the future. They each have their own advantages and disadvantages towards interaction. VR offers an immersive environment with multiple viewpoints for easy navigation; while AR provides an exocentric view blending virtual content into the real world. Could one obtain the best of both worlds? New types of rich digital content, display techniques and interaction can be developed within this research investigation.

A truly immersive experience consists of many aspects, and the essence of being "there" is of huge importance. Game and film reviewers mention immersion as being related to the realism of the virtual world or to the atmospheric sounds. Immersion is also perceived to have depth. The experience of immersion is a critical ingredient for complete

user enjoyment and can be made void by certain characteristics. It is possible to have a realistic world filled with atmospheric sounds and yet immersion is not achieved [1]. Likewise for VR, William Bricken captured its essence with this statement: “Psychology is the physics of virtual reality” (quoted in Wooley, 1992, p21). Virtual worlds are constructed by the senses and only really exists in the minds of users. VR is a medium for the extension of body and mind [2].

The ultimate aim behind our work has been explored for more than two decades, with researchers endlessly trying to recreate experiences and intuitive interfaces by blending reality and virtual reality. The ultimate goal would be for people to interact with the virtual domain as easily as with the real world. Various approaches have helped us achieve this - in the area of tangible interfaces, real objects are used as tools for interaction; in Augmented Reality, high quality 3D virtual imagery are overlayed onto the real world; and in Virtual Reality (VR), the real world is completely replaced with a computer generated environment.

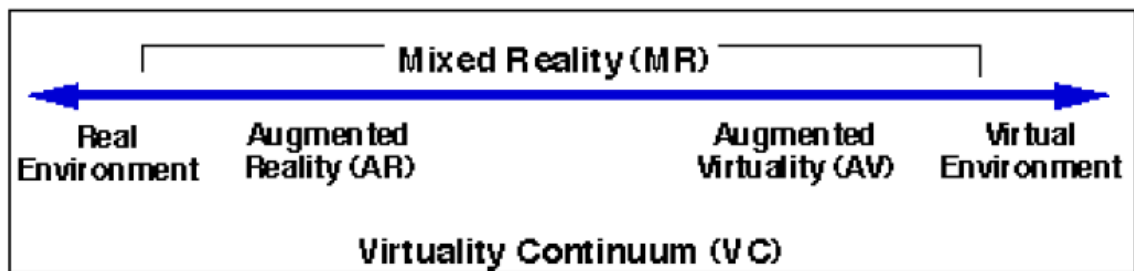


FIGURE 1.1: Simplified representation of a “virtual continuum”[3]

Milgram stated that these types of computer interfaces can be placed along a continuum according to how much of the user’s environment is computer generated [3]. Refer to figure 1.1. On this Reality-Virtuality line, immersive virtual environments lie on the far right while real tangible interfaces lie far to the left. Augmented Reality is somewhere close to the middle where the virtual is added on top of the real physical world, and Augmented Virtuality is where real content is brought into computer-generated scenes. However, human activity cannot be always broken easily into clear discrete components and users might prefer to be able to move seamlessly along this continuum as they perform their tasks [4].

What we want to also achieve in our work is to seamlessly transport the users from Reality and Virtuality, but should we or should we not let this transitioning be within their control? This is acceptable when users are highly interested in a scene and want

to be fully immersed and focused in the virtuality; however while being visually engrossed, they might feel like grabbing a drink but they could not unless they remove the head-mounted display (HMD). Another example would be collaborating and having conversations with other users, they definitely would not be able to do so while wearing the HMD. In a HMD, a person is separated from the real world, taking away their most important sensory channel - their sight in reality. Thus it is gravely important that users be able to move seamlessly between the real and virtual world. We want the computer interface to be invisible, such that the user can interact with graphical content as easily as any real object through their natural motions. For example in [5], the turning of a book page to change virtual scenes is as natural as rotating the page to see a different side of the virtual model. Holding up the AR display to the face to see an enhanced view is similar to using reading glasses or a magnifying lens. Though the graphical content might not be realistic, looking and responding naturally like real objects would increase the ease of usability. To create a believable experience, two things are dependent: interactivity and precision. These can either make or break an AR application, and they are also crucial in reducing motion sickness.

With the huge importance of visuals in providing an immersive watching experience, rendering 360-degree panoramic images/videos must be done well. Such content are increasingly available recently not only for head mounted displays but for many other display devices as well; for example Google's introduction of photo sphere for Street View¹ and YouTube's support for 360-degree videos².

3D interaction with live film content is challenging and involves addressing complex problems in computer graphics, such as being able to estimate 3D geometry, materials, lighting, and other extensive 3D scene information. This project will perform research into Reality and Virtuality transitioning whilst keeping in mind how best to expand this to include renderings of 3D virtual objects into the filmed background. Previously AR and Mixed Reality (MR) technology has both been used for live composition of real and virtual scenes, balancing the immersion from being in a virtual world while looking into the physical world for real physical interaction. Users will view the virtual content on a consumer level head mounted display (HMD) such as Oculus Rift. By combining tracking [6] and geometry of the user, the real-world user (and accompanying objects)

¹Google Street View : <https://www.google.co.nz/maps/streetview/publish/>

²Youtube 360 Videos : <http://www.theverge.com/2015/3/13/8203173/youtube-now-supports-360-degree-videos>

can be easily incorporated into the movie scene to interact with the digital content. Although HMD is an ideal device for delivering immersive experiences, it often causes significant eye fatigue and simulator sickness. Correct focusing in a scene and high frame rate rendering on a stereoscopic display can reduce visual fatigue and discomfort.

The idea of this work being a home entertainment system lead us to follow in the footsteps of regular household video game consoles which are portable and easy to set up. We, too, want this system to not require any special expensive environments, but make use of cutting-edge devices and explore compatible techniques.

The thesis is broken into several chapters. First chapter is *Background* where it summarizes and studies various past research in the related fields of image composition, augmented reality, augmented virtuality, virtual environments, and understanding towards sense of Presence.

The next chapter *Design* will describe our inspiration and concept of the whole system. It will also link the understanding of sense of Presence and immersion to how our approach was carried out. The chapter will close with a list of research questions describing our aims in this investigation.

The chapter *Implementation and Prototype* will give a description following an explanation of our system requirements and what the software application consists of. Then we show the development process of our system prototype, how certain problems leading one after another were attempted and solved. The design decisions for both hardware and software are explained and shown.

User Evaluation will cover the iterative evaluation process. Beginning with results from an initial focus group, it proceeds to describing the TEDx conference where a demo of the system prototype was shown. Results and feedback were obtained from TEDx to create the next round of changes to the system. The user experiment design is described and elaborated with reasons supporting key design decisions. The materials used and procedures carried out during the experiment are also explained. In addition, the measurement tools used for evaluating sense of Presence and level of user engagement are described in detail too.

The chapter *Results and Discussion* provides a detailed description of the results, followed by explanations and thoughts behind them. All discussions of the results are further supported with qualitative feedback and quantitative data. Statistical tests used for analysis were explained with their results linked to previously stated understandings and concepts. Explanations were drawn carefully with supporting participants feedback on the exclusion of a few of the participants from analysis. Lastly, we listed the improvements that participants wanted and hope to see working in future cinematic virtual environments.

Conclusion and Future Work will conclude the thesis with a summary of all results and new discoveries. A list of contributions by this research investigation is given along with future works for improvement.

Chapter 2

Background

Our research is based on previous work on augmented blending, augmented virtuality, 3D reconstruction, mixed reality, head-mounted displays, and immersive user experiences through the use of game engines. In this section we discuss related works from each of these areas in turn.

2.1 Computer Graphics, Game Engines

With the rapid proliferation of desktop and console gaming in all directions, game engines are not only more powerful and easy to learn but also increasingly affordable for both developers and academics alike. The features and tools that the Unity 5¹ and Unreal Engine 4² graphics powerhouses can provide are immensely more than capable to support today's demand of entertainment content. From physically-based standard shaders, cascading visual post-processing effects, highly flexible editing and visual scripting, to leading middle-ware integrations, these technology systems supports every need of this project and glues them together presenting a refined finished product at the end. Oculus VR supported integration makes these game engines even more of a must with the huge ease in respective developments. All one has to do is install the general Oculus driver, plug in the headset, switch the engine to VR mode and run play.

¹Unity 5 <https://unity3d.com/>

²Unreal Engine 4 <https://www.unrealengine.com/>

Unity 5 and Unreal Engine 4 are equal with their abundant learning resources and friendly helpful community. But we chose to use the Unity 5 engine because of these key reasons:

- here at the HIT Lab NZ there is a wealth of experiences and knowledge on knowing how to use it
- Unreal 4 just got publicly released to the community recently, thus there's a lot we do not know about it regarding its usage
- The RGB-D SoftKinetic camera that we planned to use has sample projects built in Unity thus this will speed up any development
- Building a first prototype quickly is a priority and with the Unity engine, it is possible

With demands of high quality rendering and support for leading industrial middle-ware technologies used all covered by the game engine, all that is left is getting and processing the image and depth data obtained from the attached optical camera. The user's body has to be augmented/blended harmoniously into a cinematic scene and appear as if it is part of the virtual environment. Various previous works has approached this in many different ways with diverse equipment and architecture used.

2.2 Image Composition, Augmented Reality

One example of blending the user's hand into cinematic scene would be would be [7] which provides three techniques using shaders to provide realistic hand occlusion, shadows, and visual distortion.

The first technique involves using depth data from a depth camera, and match them to a skin color segmented image, to create a depth map containing only values assumed belonging to the hands. This depth map is then passed to the shader to test the depth of each pixel at render time. If the hand pixel was closest then no virtual content is rendered at this location, and the texture of the hand is shown instead. Its advantage was its sharp per-pixel occlusion however it is important that the field of view of the

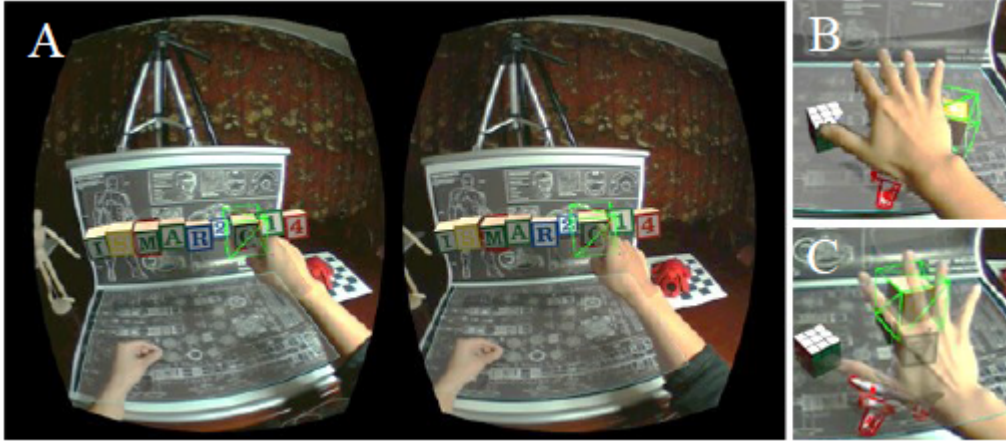


FIGURE 2.1: (A) AR-Rift's view, (B) Hand occlusion of method A and B, and (C) Hand occlusion of method C [7]

depth camera has to match as much as possible with the Oculus Rift's. A mismatch could lead to inaccuracies or occlusions will only be detected at certain regions.

The second technique is less accurate however it removes the need of a depth camera by using ray-casting from the viewing camera position to each finger of the user's hand. It could be improved through calculating a disparity map from the stereo cameras however this would spend more computational resources.

The third technique that was used is a semi-transparent hand reconstruction approach. It involves using the hand model that was reconstructed for hand shadows as a semi-transparent proxy for the user's actual hands. This allows users to see the virtual hands overlaid over their actual ones, while still having shadows cast on virtual objects. Their participants felt they could perform equally well using the third method or the previous two techniques that are more computationally expensive. The disadvantage of these techniques though is its dependency on the 3Gear Nimble system and an additional setup with a color depth camera on a tripod, pointing down on the interaction space which they used for hand tracking; and especially the third technique which requires a calibration step to align the virtual hands with the real hands as precisely as possible. Therefore modifications have to be considered with the possibility of including capabilities to perform body occlusion too if we were to incorporate this technique in our system. For our project, our main goal is not on gesture recognition which is what the 3Gear is for though the AR-Rift implementation with the wide-angle stereo camera setup is useful, and we could take this on-board. However the depth color camera, that is mounted on top of the AR-Rift, utilised for hand occlusions has a mismatch field

of view between the AR-Rift stereo cameras. This means that occlusion can only be applied for a small portion of the whole scene. Also with all these 3 cameras mounted on the HMD, it makes it feel heavy reducing the overall user experience.

For natural interaction in Augmented Reality (AR) to be widely adopted, its techniques needs to support precise interaction, and gestures have to be simple and cognitively easy for users to relate to. [7] shown that the free-hand gesture based Grasp-Shell is more preferred and efficient than the multi-modal Gesture-Speech which combines speech and gestures. However together they complement each other perfectly demonstrating good control and interactivity in a physics enabled augmented environment. In [6] where a live complete 3D model is created, for our project the model could be the user's body, this would allow composited virtual graphics to be precisely occluded by the user, including geometrically complex objects. These virtual objects can then be rendered from the same perspective of the tracked color depth Microsoft Kinect camera, enabling objects to be spatially registered as the camera moves. Hand tracking here similarly would be very easy to do. There have been extensive research conducted in this area, one common technique to track hands was to use color as a cue for object localisation, segmentation and tracking [8]. Another way that is more complicated but reliable is explained in this paper [9] where it involves training, weights, descriptors but still based on the most reliable cue for hand tracking obtained by skin color detection using classifiers applied to a color space with brightness normalisation. Refer to figure 2.2 where hand tracking is applied.

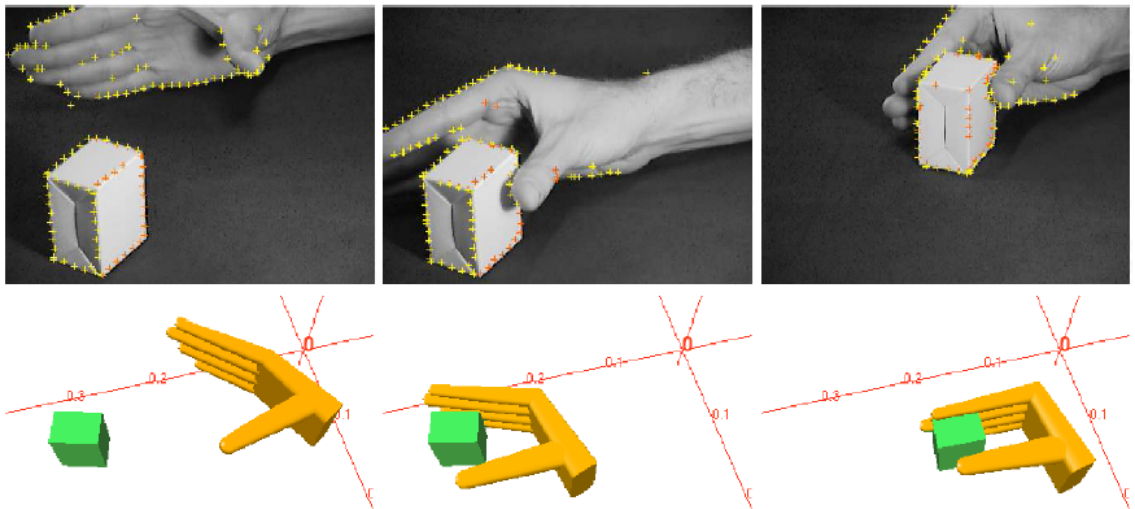


FIGURE 2.2: Tracking a hand grasping a box. The graphical images show the view from above [9]

Other methods involved using data gloves but this affects the natural look of the hands and is only capable of tracking hands [10]; using an external tracking device like the Microsoft Kinect has the potential to do a lot but it requires a precise calibration of the system and quickly leads to inaccurate segmentation results when you step out of the "sweet" zones, not to mention the amount of customization and configuration needed in each set-up [11] [12]. There is no doubt over the precision of the color chroma-keying technique in its segmentation [13][14] when the guidelines are followed strictly but it would be interesting to see if there is any way of tracking it. Also having it extended to regardless of whether the user's hands and body are covered with clothing would be very convenient. A main disadvantage of it is being restricted to a fixed type of room using the green box approach where either the background is uniformly colored with a color that ideally does not occur in the foreground body parts [14], or the environment should not contain any yellowish or reddish objects that could be confused with the skin tones of the user [13]. Refer to figures 2.3 and 2.4. In addition if the user's skin tone is darker or in a different color, this technique will most likely fail too. Therefore instead of color keying, we want to explore depth keying with a depth capable camera and see the results it produces.



FIGURE 2.3: (b) Camera images showing the user's hands in the real world and (c) the user's virtual view with the segmented hands [14]

The paper [6] demonstrates how a user holding and moving a standard Kinect camera to rapidly create detailed 3D reconstructions of an indoor scene. Utilising the depth data, they were able to track the 3D pose of the sensor and reconstruct geometrically precise 3D models of the physical scene in real-time. It is known that Kinect camera's point-based data is inherently noisy, making high level surface geometry difficult to



FIGURE 2.4: Usage of the Aughanded Virtuality application [13]

generate. To do this, their system continually tracks the 6DOF pose of the camera and fuses live depth data from the camera into a single global 3D model in real time. As the user explores the environment, more of the space is revealed and fused to the same model. With relevance to our work, we will not be able to obtain the user's body through only just one glance with the SoftKinetic DS325 camera. However by applying this we are able to form a more detailed mesh model of the user every time he/she looks at himself/herself. They have made one fundamental assumption in this reconstruction approach, which is that the scene has to be static; so that by tracking the 6DOF pose of the camera they would be able to construct a geometrically precise 3D model of that stationary scene. In our work, the DS325 is mounted onto the Oculus Rift, therefore we will be too considering the 6DOF of the camera but the user would be moving at different points in time and so their approach will not precisely work well as we would need to be able to tell which pixels are part of the user's body now and which are not. It is known that their method is robust to fast scene motions and changes but prolonged movements gave them problems.

The paper [15] is another that used a depth camera attached to an Oculus Rift. For scene reconstruction and occlusion of the user's hands, they first grab the depth map of the entire environment and repair it by removing noise and filling holes. Next they converted the RGB-D (RGB and the repaired depth map) to a mesh by first creating a 2D mesh grid of vertices of the same size as the depth map, before using a GLSL vertex shader to position each mesh vertex according to the depth values in the map.

To apply occlusion they capture and store the static environment without the user's hands present. Then during the experiment they set the camera up to capture pixels of the closest surface which would highly likely be the user's hands and arms; this is the dynamic scene. Both scenes are finally rendered after fragments in the dynamic scene with similar depth values to the static scene are discarded by the fragment shader.

2.3 Realism, Lighting, Augmented Virtuality

Apart from the accuracy and quality of reconstruction and occlusions with user hands, composited synthetic objects, and the user body, another important factor to achieve high degree of realism when watching a 360 panorama movie are the relative lighting conditions from changing environments in the cinematic scene. Superimposed objects have to be placed at exact positions where they would exist in the real world. Ideally, light sources, reflections and shadows will match in both real and virtual worlds to obtain consistent global illumination; hence movements in the two worlds must be synchronized. As rendering is done in mixed reality demands real-time performance, image-based approaches like image-based lighting and environment illumination maps are needed. The type of high quality photo-realistic scenes can be seen in figures 2.5 and 2.6.

In [14] a simple approach would be to adapt the brightness of the pixels in the camera image of the user's body to the brightness of the virtual environment. The brightness of the current view of the virtual world can be calculated by averaging the intensity values for each pixel. The image of the user can then be brightened or darkened such that the average intensity values of all pixels matches the average of the virtual environment. Image-based lighting is meant for integrating computer-generated imagery with photographs realistically using measurements of real-world lighting to illuminate computer-generated objects. Incorporating this technique would definitely influence the quality and realism of our user body visualization, enhancing the sense of presence viewers will have. However this would take time to be developed and from a research point of view, it is enough to have a simple "see-through" graphics based on the color image

³Koola UE4 https://twitter.com/Koola_UE4/media <http://bit.ly/1mQ5EUA>

⁴Koola UE4 https://twitter.com/Koola_UE4/media <http://bit.ly/1mQ5EUA>



FIGURE 2.5: Koola : Color³

FIGURE 2.6: Koola : Swamp⁴

provided from the camera to test the concept of user body visualization and observe its effects.

2.4 Head-mounted displays, Virtual Reality

A closely related work [16] explored mixed reality rendering techniques that incorporate the real physical world into the virtual world enabling user interactions with physical objects. The key is to balance immersion in the virtual world with viewing the real world for physical activity. Despite being highly immersive VR, HMDs have one flaw in occluding the real world and making social physical interactions for the user difficult and awkward. Taking the HMD off would break the immersion the user is experiencing which is proven to be time costly [1], making such context switching expensive.

Results showed that users prefer rendering which is a selective blend of virtual and real, while maintaining the one to one scaling of the physical environment [16]. Users wanted to see their hands, the object of interest and also salient edges of the surrounding environment. Users were more acceptable of viewing the real physical environment overlaid on top of the virtual, or taking the HMD off when watching a movie; but not so much when the experience was highly engaging like in the FPS and racing game. Based on these users' comments, we could see how important a feature to control the visibility of the real physical environment in correspondence with the virtual is. Therefore we learnt that giving users that control would allow them to fit the immersive experience to their own comforts and needs.

The expected launches in the coming years of several major headsets, Valve's Vive⁵, Oculus Rift's new Rift⁶, Sony's Morpheus⁷, Samsung's Gear VR⁸, Microsoft's HoloLens⁹; just to name a few, are incredibly exciting times for the industry. The possibilities of what we can do with Magic Leap¹⁰ are further enticing. Right now the mixed/virtual reality community ecosystem is still small with new 360 panorama content growing from more than just video game making but movie-making too. Being a entirely new medium for cinema, film directors and studios are curious about the possibilities of MR and VR; Oculus Rift responded with its new filmmaking venture, Story Studio¹¹ and they are not alone with many more such as Penrose Studios¹², etc. It is a revolutionary departure from the moving pictures of today into the future of story-telling in virtual reality. Let more "wow" moments be created with the user "being" in the movie instead of just watching; but one thing was missing when this journalist¹³ crouched during his immersion in Oculus Rift's Lost VR demo, his real physical body.

2.5 Virtual environments, Sense of presence, Self-avatar

Does having every object being virtual inhibit the interactivity and level of immersion? If participants spend most of their time and cognitive load on learning and adapting to interacting with virtual objects, does this reduce effectiveness of the virtual environment? This study [17] looked into whether interaction with real objects and seeing a visually faithful self-avatar improves task performance; it also explored if having this avatar does improve the sense of presence. Self-avatar refers specifically to the virtual representation of the user. Results showed that user task performance was close to real life when manipulating physical objects in a VE rather than virtual objects. There was also no significant difference in the sense of presence regardless of the self-avatar's visual fidelity or presence of real objects. Motion fidelity was proven to be more important than visual fidelity for self-avatar believability. Efforts should be focused to tracking and animation of the user rather than rendering quality of the VE.

⁵Vive www.htcvive.com/

⁶Rift <https://www.oculus.com/en-us/rift/>

⁷PlayStation VR <https://www.playstation.com/en-us/explore/playstation-vr/>

⁸Gear VR www.samsung.com/global/galaxy/wearables/gear-vr/

⁹HoloLens <http://www.microsoft.com/microsoft-hololens/en-us>

¹⁰Magic Leap www.magicleap.com

¹¹Oculus Story Studio <http://storystudio.oculus.com>

¹²Penrose Studios <http://www.penrovestudios.com/>

¹³<http://bit.ly/1BulowU>

More general research on embodiment and sense of presence in virtual worlds is explored by [18] where they focused on concepts, theories and insights regarding embodiment afforded by an avatar. Presence is a multi-dimensional cognitive process with many complex models. One interesting relation to it is with the judgement of the virtual environment's realness where the immersive virtual environment participants are viewing is identically represented by the same real environment that they are physically occupying. This set-up appears not to exhibit the same strong indications, previously observed by others, of perceiving distances to be compressed in the virtual environment [19]. If it turns out that egocentric distance compression, is, in fact, a symptom of a lack of 'presence' in an immersive VE, it may become possible in the future to use such action-based metrics to quantitatively assess the extent to which a participant experiences 'immersion' in a particular virtual environment. Another research also on space perception in the VE but with self-avatar embodiment, explores the influence of self-avatars on one type of spatial judgement requiring absolute distance perception [20]. It was found that the presence of an avatar changed the typical pattern of distance underestimation seen in many HMD-based virtual environment studies. A better sense in distance estimation was shown when the avatar was animated in correspondence to their own real body movements.

More studies conducted showed that representing the user with an virtual avatar in the environment improves the predictability of size relation and distances from an egocentric perspective [20]. Also the sense of presence increases within the computer generated scene if the user sees a body acting accordingly to his/her movements [21]. To utilise these benefits further, one could use real body video overlays rather than animated self-avatars to represent the user. This opens possibilities of bringing visual characteristics such as pigmentation and hairiness to the virtual world rather than making them up. One particular work presented an egocentric representation of the user's real body through a captured video stream into virtual immersive environments. Limitations are that this views only the user's hands and is based on a color chroma keying approach segmenting regions that correspond to skin tones therefore the real physical environment cannot contain any yellowish or reddish objects or anything that matches the skin tone of the user. However within those restrictions, the algorithm was very responsive and performant. As mentioned earlier in section 2.2, color chroma keying is highly accurate and effective but it requires a long laborious set-up. We want to be different in this

aspect and look at other approaches that would be easier for the general public.

Virtual reality research has the concept of presence defined as the "extent to which a person's cognitive and perceptual systems are tricked into believing they are somewhere other than their physical location" [22]. With relevance to computer gaming, a grounded theory was used to create a robust division of immersion into three levels: engagement, engrossment, and total immersion [1]. Game reviews often relate immersion to the realism of the world created or to the atmospheric sounds. This atmosphere is created from the same elements as computer generated environment construction. What makes atmosphere distinct from scene construction is relevance. Features presented must be relevant to actions and the location of the user. The more effort invested from the user to attend to the sounds and visuals, the more immersed the user would be. "The fact that you have to rely on your own senses. Visual. Auditory. Mental" [1].

2.6 Critique

I'm in VR!: using your own hands in a fully immersive MR system [23]

Similarities: 1) Their first hypothesis is that the introduction of photo-realistic capture of user's hands in a coherently rendered virtual scenario induces in the user a strong feeling of embodiment without the need of a virtual avatar as a proxy. This is good and is what we want to present with our project too. But our hypothesis is more general with not only the user's hands but his/her whole body, and our RGB image feed is not photo-realistic.

2) They had a general hypothesis similar to what we want to achieve too : By presenting to the users an egocentric view of the virtual environment "populated" by their own bodies, a very strong feeling of presence is developed as well.

3) Our approach are both different from previous work in the use of a single RGBD camera mounted on the user head is more similar to the work of [24] (they used the camera to see other people, not the self).

4) Depending on the form of MR cinema turns out to be, if it involves a lot of movement, their realization of a fully untethered Virtual Reality system where the user is immersed

in VR by means of a wearable computer and no cables is highly desirable. Otherwise a desktop-based station would be sufficient enough.

5) They also have the limitation of showing the fingers, hands, wrists till half of the forearm. Also this image does not cover the whole resolution of the Oculus Rift lens which is one of the limitations of our first prototype.

Differences: 1) They used a CAVE set-up for their system doing it similar to the Huge Immersive Virtual Environment (HIVE) presented in utilising wide-area tracing coupled with HMD-based immersive VR, resulting in a relatively large active space able to simulate a virtual space at a one-to-one scale [25]. While we are using a desktop set-up since we do not have any positional tracking in place for our demo application, and also this is a movie-watching cinema experience therefore we do not expect the user wanting to walk around a lot.

2) Their rendering pipeline consists of using GPU shaders (vertex, tessellation, geometry and fragment) and a scene graph to manage the real-time capturing and rendering of user's hands. While we are using the Unity 5 engine to handle our asset game objects and rendering.

Claims: At the current state of technology, accuracy of avatar representations for user interactions is still limited such that they rarely correspond exactly to the dimensions or current posture of the user without wearing any tracking markers.

WeARHand: Head-Worn, RGB-D Camera Based, Bare-Hand User Interface with Visually Enhanced Depth Perception [26]

This paper is mostly about the hand interactions/techniques to elevate performance and usability. Task would be to manipulate any kind of virtual 3D object using bare hands in a wearable augmented reality environment. In this paper, their HWD is an ACCUPIX mybud with a resolution of 852x480 (WVGA) which is small in comparison to the Oculus Rift. Also the camera they used for hand capturing is a Creative Interactive Gesture near-range RGB-D camera with a color resolution of 640x480 and a depth map resolution of 320x240. Clearly you could see that they could easily double the size of the depth map to fit the color image which covers the whole width and three-quarters the height of the head wearable display. This is something we could not do similarly with the SoftKinetic DS325 and Oculus Rift DK2 display.

Full Body Acting Rehearsal in a Networked Virtual Environment - A Case Study [27]

This paper is a good analogy to how our system would change the conventional movie making/directing. In our context, the director would be able to put on the HMD and view the actors acting from the perspective of a viewer in real-time as though there is a completed movie reel playing the visuals.

Claims: Mixed reality (MR) is not new in film making. In Tenmoku, Ichikari, Shitbata, Kimura, and Tamure (2006) [28] and in Ichikari et al. (2010) [29], the authors propose a workflow to insert computer graphics animation data (i.e. created animation in 3D animation software), motion capture data and 3D video data into special film-making software. The positions and movements of a camera can be derived from the 3D animation and incorporated in a 3D model representing the physical location of the filming. By using 3D animation data and a 3D model of the scenery, the director is able to plan in advance the movements and positions of the camera for the shot to be filmed. In Ichikari et al., a real-time method for relighting the 3D scene was also added to the previous method.

Differences: 1) They are looking at shared virtual environments where actors and the director could move virtually as their own avatar in the virtual environment, make simple gestures and change their facial expression through keyboard presses. For us, additional work is needed to be able to create spherical 360 panorama videos of real-time acting on-the-fly and stream it directly to our application so that the director could immediately view it as a spectator. Also our system extends the scope of film making as one has to consider what actions can the spectator perform or type of interaction with characters in the film.

2.7 Conclusion

In this section we have seen work touching base on every aspect of our target system. With the target of a regular home entertainment system, we have to not only consider what can be done but also what it should be. No matter which presence model or theory observed, the user body visualization is known to affect presence in some or many ways. Not many work has looked into the virtual environment being a 360 movie played in real-time, enveloping the viewer into the virtual cinematic world and its effects of using

the techniques mentioned above to display the user's real body. A considerable amount of work has explored the effects on presence by giving the user a virtual avatar, now we want to look into a real body visualization instead. The depth keying approach was used rather than color chroma keying because of chroma-keying's strict guidelines and inflexibility to change environments and huge training required in order to be robust to color and hue variations. Though lighting, reflections, shadows etc are crucial to the realistic quality of the user body visualization, our research goals could be met through using simple graphics and have the time better spent on other system considerations. Also with a real-time played 360 movie as the background of the virtual world, there is no simple solution to render a realistic body visualization.

Chapter 3

Design

3.1 Inspiration and Idea Generation

This chapter describes how the inspiration for the AR experience came about, and the planning involved for it to be delivered on a wide field-of-view head-mounted display called the Oculus Rift, with a depth and RGB camera similar to the Kinect camera, called the SoftKinetic camera. This camera is mounted onto the Oculus Rift providing the technology for body augmentation. Development included a software program capable of capturing and compositing the user's body into the cinematic scene with keying technology. There is a number of things to take note of unique to AR, such as the need to provide virtual depth cues in the real world and occlusion of the user's hands.

The general design goals that we set out to achieve are: 1) Hardware setup and keying based user's body augmentation 2) System integration and user evaluation

3.1.1 Brainstorm

The research aim is to blend the user's body and environment into a cinematic scene and investigate its impact on immersive user experience. A prototype system was targeted to be developed to capture the real-world user and accompanying environment and blend them into a cinematic scene in real time. Also a basic mechanism will be built performing automatic transitions of the user's body in and out of the scene. This could be further expanded to perform controlled transitions based on certain user input



FIGURE 3.1: Concept Scenario

parameters coming from an interface. In figures 3.1 and 3.2, you could see the viewer in his real world surroundings and able to see himself. Clockwise from the top left onwards in figure 3.2 we see the viewer fading from the physical real world to the cinematic virtual world. First we fade out the viewer's far surroundings, followed by blending out his near surroundings, then items beside him will disappear too before his body also fades away, leaving the user in the virtual with no representation of himself or his physical surroundings.

Immersion is used to describe the degree of involvement with a certain media. This involvement moves along the path of time and are hit with barriers. Some barriers can only be removed by human activity such as concentration; others can only be opened by the virtual environment itself such as scene construction. Each level of involvement is only possible if the respective barriers are removed. Even removing these barriers only allows for the experience and does not guarantee it. Three levels of involvement were found: engagement; moving on to greater involvement in engrossment; and finally total immersion [1].

Even practitioners familiar with VR would be confused by the terms immersion and presence, which define distinct concepts. Mel Slater defines them this way [30]

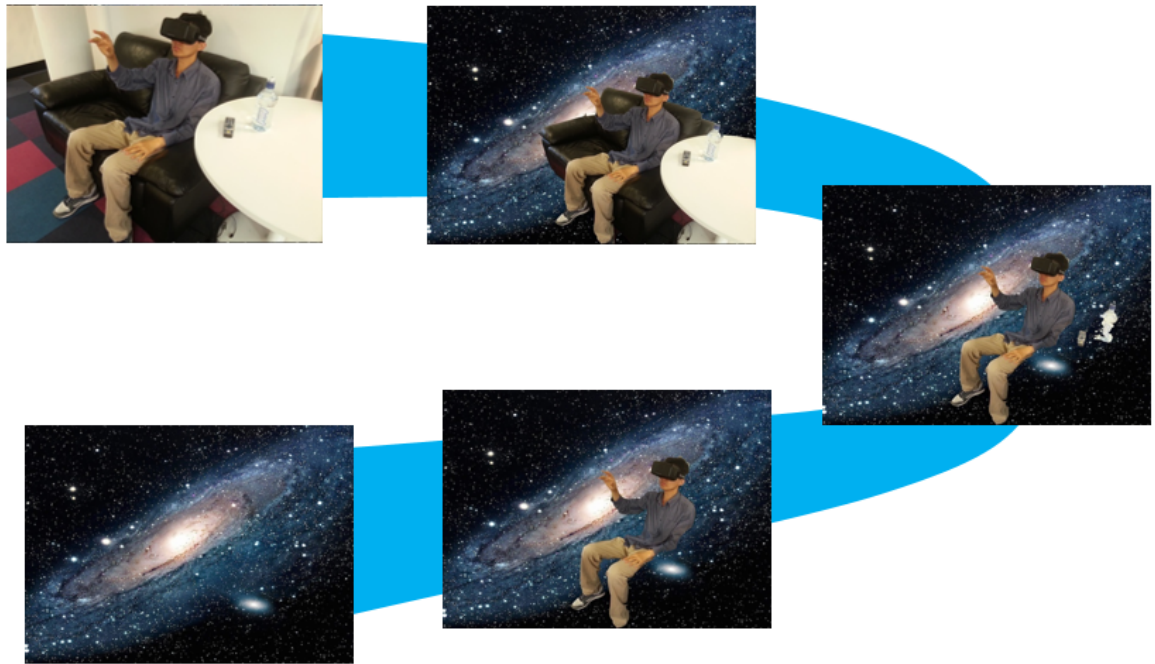


FIGURE 3.2: Concept Idea - Transitions between Reality and Virtuality

- Immersion refers to the objective level of sensory fidelity a VR system provides
- Presence refers to a user's subjective psychological response to a VR system.

A VR system's level of immersion is dependent only on the system's rendering software and display technology including all types of sensory displays. Immersion is objective and measurable—one system can have a higher level of immersion than another. In this element, we reckon our system can only be pushed so far. After all, what we intend to research on is not on our ability to produce amazing graphics. What we want is to look into if viewers feel more "being there" in the virtual world when they can see their real body visualization. And with their bodies being transitioned into and out between virtual and reality, we were interested in how would this experience variable vary. In this case, presence is a better fit to test for since it is an individual and context-dependent user response. Different users can experience different levels of presence with the same VR system, and a single user might experience different levels of presence with the same system at different times, depending on state of mind and other factors [31].

Atmosphere is created from the same elements as computer generated environment construction. The graphics and sounds combined create this presence. What makes atmosphere distinct from scene construction is relevance. The features must be relevant to the actions and location of the user. The reason this is important is because of the use of attention. The more effort invested from the user to attend to the sounds and visuals, the more immersed the user would be. "The fact that you have to rely on your own senses." [1] This demands attention and more specifically the three elements of attention: visual, auditory and mental. Though what we aim to do in this work does not involve sound, by tuning the creation of a virtual cinematic world we would like to test how much would the user be engaged by visuals alone.

Ultimately like what Pixar did with Toy Story, we want this technology to be used in the right way, telling the right "story" [32]. When Pixar did their rendering, to them everything looked like plastic. That led them to think "What if the characters were made of plastic? What if they were toys?". This shows how important the content is and the feel it gives off to viewers. As Pixar focused on the story and hid technology, we similarly want people to be pulled into our content, our story and the features that it offers, and not think about the technology. After all it is not the technology that entertains people, it is what you do with the technology. It is important to make it invisible but still have it push to do something new.

3.1.2 User and Use case

Immersive VR has seen some success in the entertainment industry where these attraction places visitors inside the game world even more compelling than first-person games on desktop PCs and console gaming systems. While the experience can be impressive, it is limited to few installations and short gameplay times per user due to systems' high cost. Though not really a main goal, we do hope that our system can be portable and easily set-up by anyone in any regular household, and still able to deliver immersive experiences. Just like well-known home video game consoles such as the PlayStation and Xbox series, our system would be as commonly known as a form of home entertainment in the future. With our current set-up idea, we believe this was definitely possible.

Traditionally, videos and similarly movies are locked to the angle where the camera was pointing to during the filming, and the resulting video has boundaries. Now with 360°

video recording, there are no longer such boundaries. Also the general public are starting to be able to afford these 360° video capturing devices as they become more common. Video is a very rich media type, presenting huge amounts of information that change along time. 360° video features even more information in the same time, and adds extra challenges when we cannot watch all around at the same time. But it provides the whole picture all around the viewer, holding the potential to provide full immersive user experiences [33].

Our target audience and consumers would be the general public, practically anyone of mostly all ages. With the abundance of 360 video goodies greatly increasing during this time with YouTube and Facebook strong support, it is a sure thing it will be a common media entertainment format for our system as computer games are for home game consoles. Just as now you can purchase and download games from the internet to our home consoles, so you can too with our system since tentatively it will have its own designated desktop computer or you could even set it up with your home desktop.

3.1.3 Research Questions

There are a number of research questions that we wanted answered at the end of this research. In this section I will describe some of the questions and what attempts were made to address them.

- Will we improve the user experience of watching movies to experiencing them through blending the real-world user and the accompanying environment into the movie scene in real time?

It is known from our literature background, bringing the user's body to the virtual world that they are viewing will increase the sense of presence. Users shared that when seeing their own body in the virtual environment, they perceived that they are now part of it. One way of blending would be using the SoftKinetic depth camera which would stream frames of live video and we could check each pixel's depth value if it exceeds a certain threshold; if it does then we would set it not to be rendered. Controlling the visibility of the real physical environment in correspondence with the virtual world is important for high user engagement, so how about we make this a feature to users? Research found that users prefer

rendering which selectively blends the virtual and physical while maintaining the one-to-one scaling of the physical environment [16] but now with having the control of fading in-and-out at the user's will, how would their preferences change? Would users be annoyed if at certain scenes in the movie, it automatically fades? Movie directors who make 360 panorama films might want the control over the blending to maximise users immersion. Even the type of interface or gesture one has to perform to vary the blending opens up another possibility to increase immersion.

- What is the best way for users to control their transitioning between real and virtual environments?

As mentioned above, if we are not allowing user transitions to be done automatically, then an interface will be needed for users to provide their input. What kind of interface should be used? Voice commands? Hands? Dynamically controlled? Or something simple such as a mouse wheel scroll? [34] has explored this problem in depth but they have a slightly different aspect. We are similar in many ways such as being concerned about how participants' sense of presence and awareness vary during the transitioning, also usability issues around performing the transitions. Both have the VR and AR context however we do not change viewpoints from egocentric to exocentric and vice-versa like they do, we always stick to an egocentric view. The type of interface they used for transitions was the MagicLens interface (mouse ball + ARToolkit marker)[34]. It is possible that the kind of interface could influence the sense of presence, this will be thoroughly researched into.

- Will the sense of presence and level of user engagement vary from the user not seeing himself to being able to see his body and the environment?

From [35], we see that users were unaware of the presence of a three dimensional human nose model right in front of them in their Oculus Rift displays during their simulation experience. The nose was shown in simulations varying from short durations with intense motions to long durations with mild movements. Though the nose is constantly presented to the user and the body is only shown when the user looks at himself/herself, the concept that the brain and sensory system are accustomed to using that object to tune out is the same. It affirms other research that seeing a moving self-avatar according to his/her body helps

reduce simulator sickness. But how does the sense of presence vary? With lesser sickness, does that indicate a greater tendency to be immersed and engaged in the virtual environment? We are also curious to see if the sense of presence and user engagement will shift between seeing oneself and not seeing it. How would this be affected from the appearance/disappearance of the user's real physical surroundings? Would they be put off by it since it is a reminder of the real world? Also since the user would be interacting with an interface to perform body and surroundings transitions, would this cause a difference in their level of engagement? We roughly guessed that if users were to use the interface, it will cause them to shift focus from the movie to control the blending, resulting them to lose engagement with the movie.

Chapter 4

Implementation and Prototype

4.1 The Hardware Requirements

For the main design of the hardware, we wanted to emphasize on simplicity and performance. As we are going for the depth-keying approach rather than color chroma keying, we would definitely need a RGB-D camera to provide the color image feed and the depth map for performing occlusions, segmentations and many more. Also we will need a head-mounted display such as the Oculus Rift to provide the first-person perspective which will be the doorway to an immersive experience. We hoped for a set-up that is light-weight, portable, hands-free and not a hassle to put everything together. Therefore we will not be seeking to use multiple depth cameras as it might make the headset too heavy [7], or make things more complicated in terms of arrangement and different interaction spaces [11] [12]. Lastly of course, we need a pair of decent headphones to provide the audio for the movie. Audio is a huge factor in creating immersion and can be further improved through point sources and other various techniques, but for our project we would be using normal monophonic sound.

4.1.1 Time-of-flight RGB-D camera

SoftKinetic DepthSense 325 camera¹, is a cost-effective 3D time-of-flight camera capable of enabling interactions at both short and long range by using your hands and fingers.

¹SoftKinetic DS325 <http://www.softkinetic.com/Products/DepthSenseCameras>



FIGURE 4.1: Hardware set-up

This is done all without touching a screen, key board, trackball or mouse. The camera supports depth (3D), and high definition 2D color, allowing for reliable real-time streaming capture, finger and hand tracking. At this point in time, the DS325 is one of the best lightweight compact performant depth cameras on market, well known for its versatile short and long range capabilities. It is capable of producing color resolutions up to 720p HD with a field of view of 63.2 by 49.3 by 75.2 (Horizontal, Vertical, Diagonal), and depth resolutions up to QVGA (320x240) with a field of view of 74 by 58 by 87 (Horizontal, Vertical, Diagonal). Both far smaller than the technical specifications of the Oculus Rift DK2 head-mounted display. The largest resolution depth map possible would be 512x424 using the Microsoft Kinect 2, but it did not fit the criteria for our targeted product design. The Kinect device would require more work to set-up and is not as portable though its capabilities better than the DS325.

As we would be working at short-range, its frame-rate is threshold at 30 fps. But its small and compact size allows us to mount it easily on top of the Oculus Rift DK2. This was done by the mount model provided at www.reachvr.com where SoftKinetic's DepthSense technologies provides a vast pool of resources to experience VR using their depth sensors, hand-tracking middleware and interactive samples. The DS325 camera fits

the design we had in mind with the Oculus Rift DK2 quite well. In addition with the VR support from SoftKinetic, it was a logical choice to take this option.

4.1.2 Head-mounted Display

The Oculus Development Kit 2 is the latest development kit for Rift that allows developers to access amazing applications and experiences. At our lab, we have the Oculus Development Kit 2 and 1, and Oculus is the pioneer of head-mounted displays thus the best choice was to use the DK2 for our application.

4.1.3 Desktop PC

Back then early 2015, the recommended computer hardware requirements for the Oculus SDK were a 2.0+ GHz processor, 2GB RAM, and a Direct3D10 or OpenGL3 compatible video card. A good benchmark then was to run Unreal Engine 3 and Unity 4 at 60 frames per second (FPS) with vertical sync and stereo 3D enabled. The ideal demo then was full-scene 75 FPS VR rendering with stereo and distortion.

Now this recommendations have far exceeded all these, showing how much VR has grown as with the computer power of our processors and video cards. Requirements for the Rift head-mounted display are now a Intel i5-4590 3.3GHz or greater processor, 8GB RAM, a NVIDIA GTX970 or AMD R9 290 or greater video card. This has become the standard that Oculus are even offering Oculus ready PCs done in partnership with ASUS, DELL and Alienware.

4.2 The Software

VR is an immersive medium. It creates the sensation of being entirely transported into the virtual world, providing a far more visceral experience than what any screen-based media can. Enabling the mind's continual suspension of disbelief requires particular attention to detail.

Please remember that VR development at this time early 2015 was beginning to take off hugely with all the hype and big decisions movements by major tech companies.

Guidelines, recommended getting started tools and documents were constantly changing and updated to the latest. 2015 was a year where we saw huge changes in the Oculus API, SDK, Runtimes, and libraries. Similarly state-in-the-art game engines such as Unity and Unreal Engine did their share in the race to keep their systems and tools on par with developer's expectations and demands for VR development. The Oculus Rift DK2 was Oculus's latest development kit for developers to build Virtual Reality applications.

4.2.1 Unity 5

Unity is a development platform loved throughout the gaming industry for its depth of quality optimizations, speed and efficacy of its workflows; enabling Unity users to quickly create high-end multi-platform 2D and 3D applications and interactive experiences. There are so many platforms out there now and is still growing. With Unity, we are able to build our content once and deploy at a click across all major mobile, desktop, and console platforms including the web.

Unity VR provides a base API and featureset with the goal of forward compatibility for devices and software. All VR and AR support is fully integrated in their pipeline, allowing head tracking and the appropriate field of view applied to the camera automatically and rendered to the Oculus DK2 device. It is not required to have two virtual cameras for stereoscopic displays. Any camera that has no render texture is automatically rendered in stereo to the Oculus Rift. View and projection matrices are adjusted to account for field of view and head tracking. Unity's renowned play-testing capabilities now also extends to supported VR hardware. Giving us, developers a lot of power for rapid iterative editing with little effort. We are able to view the platform-specific final build and instantly made changes to preview the difference.

4.2.2 SoftKinetic iisuTM

iisuTM is a middleware that enables communication between depth-sensing cameras and end-user applications. It allows developers to add full human body interaction, and provide advanced gesture recognition functionalities. It supports powerful legacy cameras such as the original Kinect, the Asus Xtion, the Mesa SR 4000, the Panasonic

D-Imager and of course the SoftKinetic DepthSense 311 and 325 (aka Creative Senz3D). iisu supports Flash, Unity, C# and C++ environments too.

iisuTM technology analyses 3D points that could be supplied by depth-sensing or '3D' cameras. 3D cameras are cameras that are able to sample the geometry of a scene and give an image in which each pixel supplies a measure of depth of the area covered by that pixel. The resulting image, called the "depth map", is then used to generate a cloud of 3D points (x,y,z) analysed by iisuTM in its engine. Two main technologies exist to calculate this depth: Time of flight and Structured light.

DepthSense 3D cameras such as the DS325 uses Time of flight technology. This works through LEDs embedded in the camera emitting infrared light. This light bounces off objects in the scene, returns to the camera lens and then passes to the sensor. Phase comparisons between the emitted and received infrared light waves allows calculations of depth information. The amplitude of these measurements provides an indication of the certainty of the depth information, called the confidence.

In addition to depth, some cameras also supply a color stream coming from a second sensor placed beside the depth map sensor. The color and the depth sensors have slightly different optical paths, so there is a slight difference between the viewing point of the two produced pictures, which is called parallax. A process called Registration is therefore used to generate this correspondence. For Time-of-flight-based technology, it supplies a supplementary stream of UV texture coordinates giving a position for each point of the depth map in the RGB image. The registered color image will match perfectly onto the depth image. Values of this UV map go from 0 to 1. Note that negative values on the UV map mean no matching pixel was found between the depth and color map. It is found that pixels from 1.5 metres onwards away from the depth camera have negative values in the UV map. In the next section on "Improvements and Optimisations", I built my own depth-to-color mapping so that pixels with 1.5 metres and more depth can be mapped.

The iisuTM engine is made up of a set of layers as shown in figure 4.2. Each layer is in charge of a particular processing function, and sends its results to its 'children'.

The SOURCE layer is the top layer of the processing, feeding the iisuTM engine with data provided by 3D cameras. This layer manages the color, confidence, and depth

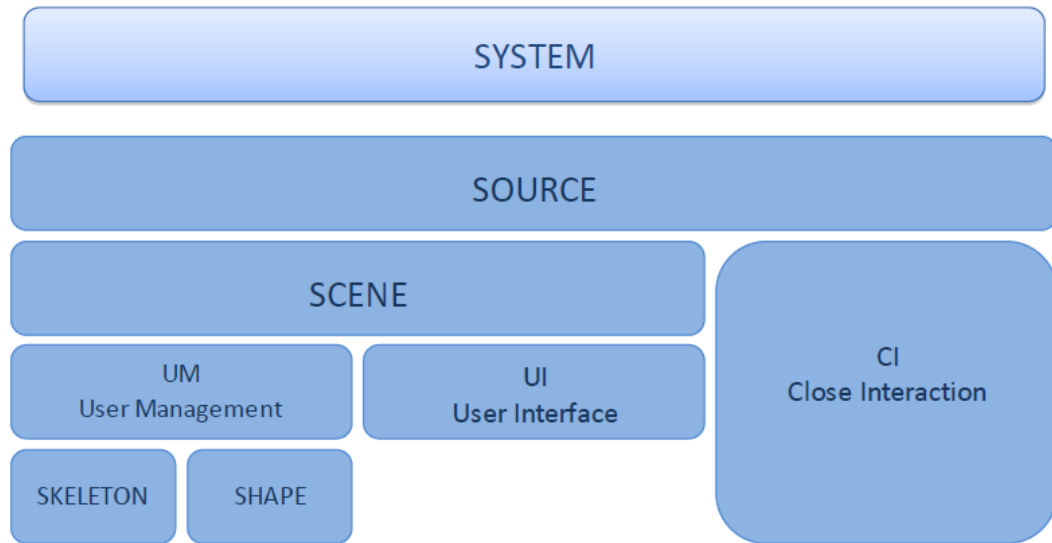


FIGURE 4.2: iisu™ engine layers

streams, offers some filter options improving camera signal for further processing.

The CI layer as shown in figure 4.3 is to detect and track hand gestures performed near the camera. CI is organized into three layers. The pipeline is illustrated with arrows to indicate data flow.

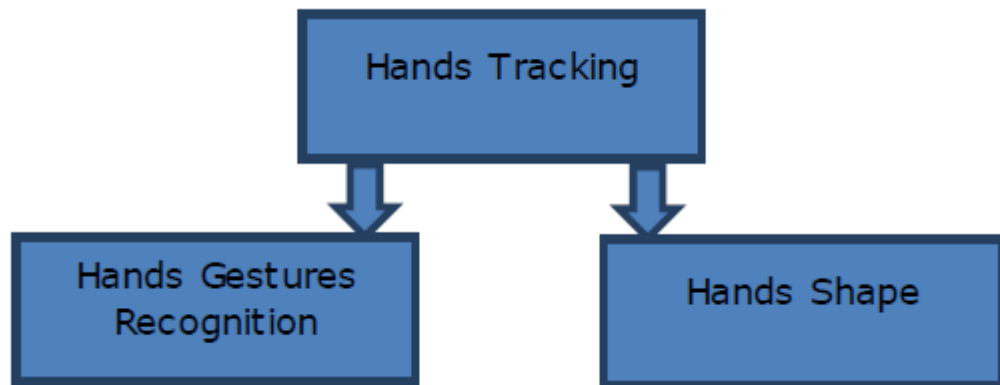


FIGURE 4.3: iisu™ CI Layers

The Hands Tracking layer detects and tracks the user's hands. It is always active when CI is active.

The Hand Gestures Recognition layer recognizes a predefined set of hand gestures. The recognized hand gestures include hand poses performed with one hand. It can be deactivated both at starttime and at runtime. I do not use the Hand Shape layer as it is not relevant to achieve the goals of the experiment.

CI automatically detects up to two hands in the camera's field of view (FOV). It detects and tracks hands that are inside the distance bounds from 0.15 to 1.0 metres from the camera. CI starts to track a hand as soon as it enters the activation zone, which is in fact a maximal distance from the camera (about 40cm). The hand has to be not touching any other surface and not occluded by another surface.

Gestures are classified into two kinds: posing gestures and moving gestures. Both kind of gesture can involve either one or two hands. iisu provides the mapping from a particular gesture, identified by a conventional character string, to its ID in current iisuTM run. The mapping between gestures IDs and names is provided as an EnumMapper associated with the meta-information of the gesture event classes. Refer to figures 4.4 and 4.5.

4.2.3 Others

EmguCV² is a cross platform .Net wrapper to the OpenCV image processing library. Allowing OpenCV functions to be called from .NET compatible languages such as C#, which is what our Unity development is mostly done in. The only problem with this was that we cannot edit any of its dlls and thus the functionalities it provides so its use is limited in that sense. EmguCV was mostly used in our work for the depth-to-color mapping calculations as it involves a lot of intensive mathematical matrix operations. There were some functions provided within Unity's own self-built libraries, but they are complicated to implement and did little of what we wanted therefore we utilised EmguCV.

There were definitely better ways to perform these mathematical calculations through other plugins, but most of them would need to be paid for. The Unity community was helpful with providing tips and directions through someone who managed to get EmguCV working with Unity's editor. This was a relief it would hold up a lot of computing resources if done manually without any optimization.

As we will be playing a 360 spherical movie in our Unity application, we need some way of rendering the video file. One way would be to use movie textures, which are animated textures in Unity, created from a video. The way how it works is when the video file is added to the project, it will automatically be imported and converted to Ogg Theora

²EmguCV www.emgu.com


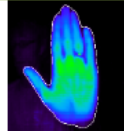




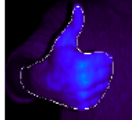
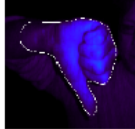
POSING GESTURES		Name
Not a recognized gesture		"UNKNOWN"
		"BIG_5"
		"STOP"
		"FINGERS4"
		"FIRST_3"
		"V_LETTER"
		"PHONE"
		"THUMB_UP"
		"THUMB_DOWN"

FIGURE 4.4: iisuTM CI Posing Gestures

"WAVE": Multiple lateral movements.

"SWIPE_RIGHT": Right interaction with a virtual plane at 30 cm. (see below)

"SWIPE_LEFT": Left interaction with a virtual plane at 30 cm.(see below)

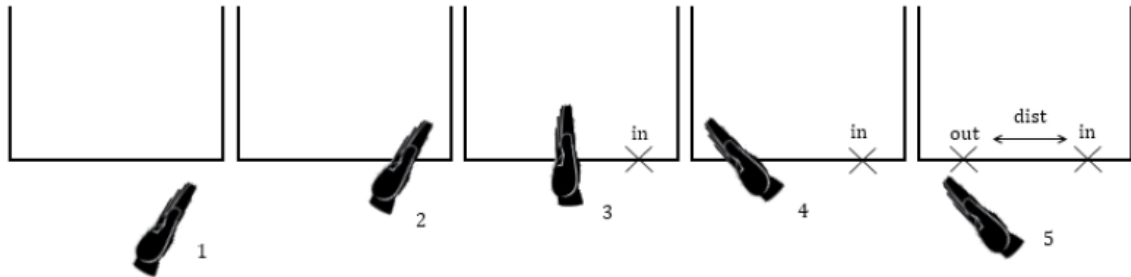


FIGURE 4.5: iisuTM CI Moving Gestures

format. Afterwards one would be able to attach it to any `GameObject` or `Material` just like any regular texture in Unity. It works but the movie texture's playback was found poor and incapable of supporting high quality content.

For an immersive experience, high movie quality is important and the playback has to be smooth with at least 60 FPS. To do this directly using shaders would be a lot of work with parallelization needed too. To get it working I resorted to the Unity community once again, and AVPro³ was discovered. Though it is a paid plugin, the trial worked exactly like the paid but it has its watermark over the playback content. It was a worthy trade off in terms of time and resources, with a slight little disturbance to the viewer's experience but not much. AVPro delivers fast, smooth, high quality HD playback through their own implemented pipeline. They only support up till Unity 5.1 thus Unity 5.1.3p2 is what the final applications is developed in.

4.3 Prototype System Implementation

The first prototype consisted of getting Unity to communicate with the SoftKinetic DS325 camera, receiving color and depth input and putting it on the display of the Oculus Rift. We also built a technique within Unity capable of watching full 360 spherical panorama videos that surrounds the viewer. Many of the 360 videos in the public media such as YouTube are called spherical videos, a type of equi-rectangular video mapping sequence. To properly view this format of videos, we have to map it on to a sphere

³Renderheads : AVPro Windows Media <http://renderheads.com/product/av-pro-windows-media/>



FIGURE 4.6: System screenshot

primitive object in our Unity scene. A sphere already has the vertices and texture coordinates perfect for an equi-rectangular projection. The only problem now is that traditionally the mapping is done on the outer surface of the sphere. In order to invert this and have it mapped on the insides of the sphere, we need to use a shader. So a basic shader is used to cull the front faces of the sphere and only show the back faces, that is the sphere's insides, in order for the mapping to be applied on the right areas. For cubic video sequences which is what Facebook 360 videos are played as, a box primitive can be used instead and it would roughly look similar.

Now that we have our video playback, we need to include body visualization. SoftKinetic support has been very helpful in this aspect; on their website, they have already uploaded some Unity project samples ranging from VR with the Oculus Rift to tracking hands and fingers to drawing the meshes of the hands. Though they have not been updated for a long time, thankfully Unity easily does this for us to make sure we are developing within the latest Unity environment. With SoftKinetic DS325 camera capability for depth perception, allowing us to get direct x,y,z coordinates of points and surfaces in the environment, we are able to yield point cloud data of the viewable space. Particle system based point cloud rendering was chosen as the technique for our visualization because of a number of reasons. Most importantly, particle system is well supported on the Unity 5 engine with many games and applications known to use it to simulate fluid

moving effects like liquids, smoke, clouds, flames, etc. Our physical bodies are moving 3D objects, therefore what could be better usage of the depth information obtained from the DS325 than to put it in a particle system. Also it seemed like a good matching representation to have each pixel in our UV and depth-to-color maps to be determined by a single particle, creating the impression of a complete physical entity/body when put together. This turned out looking really well with the point-cloud creating a depth perception that you are actually present in the virtual world physically. It was definitely better than any two dimensional image overlay. Figure 4.6 presents a screenshot showing the first-person perspective of the viewer with user body particle-system-based-point-cloud visualization enabled and a 360 movie playing in real time on the background.

With these starting tasks, I was able to put something together roughly allowing the viewer to watch a simple video clip while being able to see his/her hands and body with a first-person perspective. However there were some problems such as previously mentioned before with the poor playback framerate and the limited 1.5 metres range of the depth map. The playback issue was fixed through incorporating the AVPro trial plugin. Solving these problems would take a lot of time and require expertise in multimedia content processing. On the other hand, the short 1.5 metre range of the depth map is constrained by the hardware built into the DS325 camera, however pixels in the depth map with depth values greater than 1.5 metres can be mapped using a technique described in the next subsection.

Figure 4.7 presents an overall system block diagram, underlining each key input and output devices, bolding each software component utilised in the system, visually describing the whole architecture of our prototype.

4.3.1 **Extend UV map capabilities by building own depth-to-color mapping**

The UV map is basically a mapping between the registered color map/image with the depth map/image. Values of this UV map range between 0 to 1. The resolution size of the depth map and UV map is the same 320 by 240, giving a 1 to 1 relation of the pixels. The color map on the other hand is 1280 x 720 in size. For each pixel in the UV map, it has two values U and V which are its texture coordinates in the color map. In other words, U and V corresponds to the ratio values in the x direction and y direction

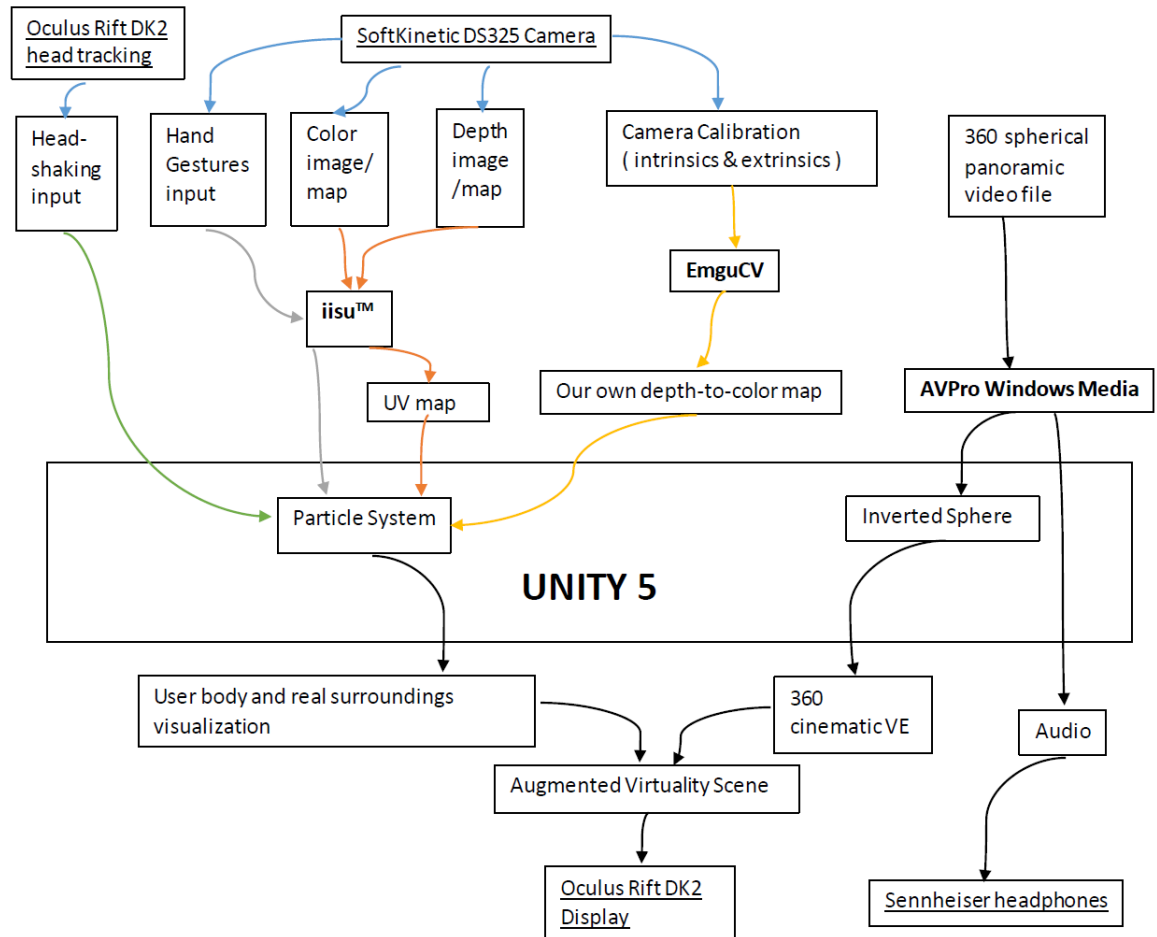


FIGURE 4.7: System Prototype Architecture

of the color map. To visually find a depth map pixel's corresponding color map pixel, one multiplies U with the width of the color map, and V with the length of the color map. However in the depth map, pixels that were found to have depth values greater than 1.5 metres would have negative values as their U and V texture coordinates. This means that the in-built registration process could only provide accurate mapping up to that accuracy. Anything beyond that 1.5 metres range we would have to map our own, which is what we did.

First we would have to perform calibration between the depth and color cameras of the DS325. As mentioned in section 4.2.3, because the color and the depth sensors are placed side by side in the DS325, they have slightly different optical paths. Thus a slight difference between the viewing points of the two produced pictures, which is called parallax. Because of parallax, camera calibration is required. The color and depth node are each modeled by its own camera intrinsics which consist of its pinhole camera model and distortion coefficients to correct the distortion of lenses. The DS325

camera extrinsics are used to determine the relative positions of the camera and depth cameras in the system - they include a rotation matrix and a translation vector.



FIGURE 4.8: SoftKinetic DS325⁴- Left lense node is for depth and right is for color

The first step to produce a depth to color mapping is to undistort the color and depth maps/images using the distortion coefficients. The estimated distortion coefficients can be obtained through the SoftKinetic DepthSense SDK once a connection with the DS325 camera is established. Then using the depth camera intrinsics, each pixel of the depth map/image can be projected to its 3D point in world depth space. To convert a 3D point in world depth space to world color space, we have to use the camera extrinsics and multiply them in. Finally we could back project the 3D point in world color space to its corresponding color map/image pixel. This is how we grab the depth map/image pixel's corresponding color for body visualization.

To make the the visualization more believable, we made use of every depth map/image pixel's depth distance value to convert this 2D to 3D by rendering them all using a particle system. Remember we only perform this costly operation for pixels in the depth map/image with depth distance values of 1.5 metres and above. Also as the depth camera has a slightly wider field of view, some depth pixels would have no corresponding color pixels. The visualization resolution is 320 by 240 exactly like the UV map and depth map; while depth pixels with depth 1.5 metres and more would be from a color map/image of resolution 640 by 480 downscaled from 1280 by 720 so that it is a multiple of the UV and depth maps' resolutions.

The UV registration done by the SoftKinetic camera DS325 is not very accurate; if one were to position his hand up-close to the camera and slowly move it further away but still in sight, you would notice the shifting of the hand skin color pixels appearing altogether rather strange. For this I had to do some guess and check with multiple repetitions to program controls over this color pixel shifting so that the visualization would look slightly more realistic to viewers; especially when they are wearing the Oculus Rift DK2 which makes pixel shifting very obvious since they are viewing up-close.

⁴SoftKinetic DS325 <http://www.softkinetic.com/Products/DepthSenseCameras>

It was realised through trial and error, user-testings and self-attempts that the most comfortable gaze when wearing the HMD is looking slightly below the middle of the screen. Thus the reason why the body and surroundings visualization is placed in the middle-below part of the display screen. No one has complained why was it placed there but there were many concerns on the small size of the “window” into reality.

4.3.2 Changing viewable depth through head-shaking and gestures

One of the original inspiring ideas that started this work was the analogy of a dream where the dream is the virtual world. As the viewer wakes from the dream, his/her reality will slowly fade in as the virtual world fades out and disappears, leaving the viewer in the real world now. One of the basic ideas to wake yourself up was to shake your head from left to right, and this fits in nicely with the Oculus Rift head-mounted display which has a gyroscope to detect rotations and an accelerometer to measure accelerations of the shakes.

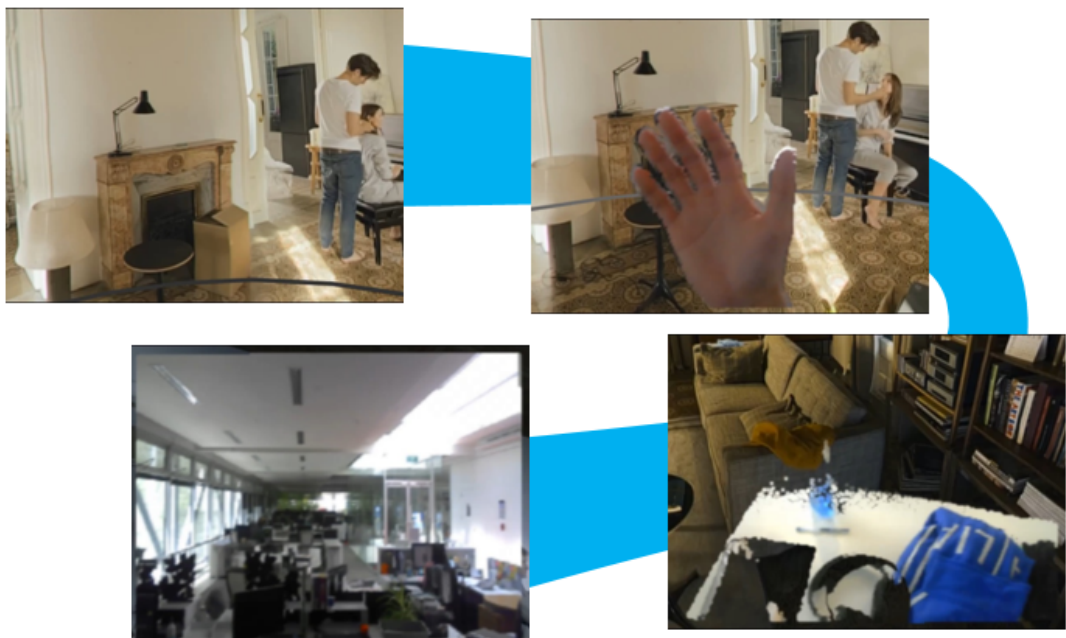


FIGURE 4.9: Waking from a dream - Head-shaking feature⁵

Our first prototype was done mainly using the Oculus gyroscope which has its values associated with the Unity environment. As the Oculus Rift integrates into Unity as the

⁵MINI Connected: Real Memories trailer <http://bit.ly/MINI--360>

main camera object which is the main camera view into the environment, it has X, Y, Z axis components and quaternions to measure its angles of rotations. Utilising its Y quaternion measurements, we can measure the variations in head movements and using timers, we can obtain its accelerations. It was working fine but requires some learning in how to "shake your head properly" because of its calculation inaccuracies sometimes. We had a few people try it out and some people were able to get it quickly while some had a hard time. Figure 4.9 demonstrates how this feature works starting from the top left and going clockwise; the first image shows the user's perspective in the virtual world only, followed by with user body visualization such that the viewer could see his/her hands, then real world surroundings visualization where the desk and objects on it can be seen, before finally seeing only the real world with the virtual all disappeared. Many people found the head-shaking dis-orientating after doing it many times which increases the chances of dizziness and nausea. Therefore we opted for an easier interaction, one which does not involve the head at all, hand gestures.

As mentioned in section 4 about SoftKinetic iisu, its CI layer is used to detect and track hand gestures performed near the camera. Since the DS325 camera will be mounted on top of the Oculus Rift DK2, it is logical to use the "Near" mode of iisu instead of "Far" mode. Note that the CI layer will not be operational if "Far" mode is selected. Each of the registered posing and moving gestures are tested for their responsiveness and accuracies. At this stage we do not want which hand is being used to matter what gestures will be detected. The more important detail would be which pose or movement is performed by the user. We found the "BIG 5", "FINGERS4", "FIRST 3", "V LETTER", "THUMB UP", "THUMB DOWN" to be more recognizable to the camera than other poses, refer to figure 4.4. And the "WAVE" is easier than "SWIPE LEFT" and "SWIPE RIGHT" among the moving gestures, refer to figure 4.5.

The gestures we used were "BIG 5", "THUMB UP", "THUMB DOWN", and "WAVE". For our system they respectively mean "stop any change in the viewable depth distance", "increase the viewable depth distance", "decrease the viewable depth distance", and "toggle the showing of HUD". Compared to head-shaking, with hand gestures there are more points of usable interactions that are easy to use and for the system to detect so it is a win situation.

4.3.3 Attempts at parallel processing

With AVPro plugin doing the movie playback, our system's frame-rate reaches above 80 frames per second. However when we include the load of our particle system visualization of the real world, it drops till about 20-25 frames per second. The Unity Profiler is perfect for this type of scenario. It reports how much time is spent in the various areas of our system. With Unity's Editor playback ability, we can run our system and watch the profiling record performance data as time passes. But because profiling requires instrumentation of our system code, it was best to do parts of the system at a time to avoid over-loading the profiler.

The profiler window displays all recorded data in a time-line with a history of several hundred frames. The time-line consists of a few areas : CPU Usage, Rendering and Memory. Bulk of our system performance was provided by CPU Usage; by clicking on the different sections, we could see each contribution to the CPU chart in detail and executing the scripts was found to be the most costliest. With the GPU Area doing almost nothing, it would be a waste to not use it for some data calculations. General Processing GPU (GPGPU) was the solution; and Unity offers these massive parallel algorithms in the form of computer shaders, programs that run on graphics card, outside of the rendering pipeline. Compute shaders in Unity closely match DirectX 11 DirectCompute technology, which is something I am unfamiliar with.

Getting DirectCompute onboard our system's pipeline was complex, and only worked with part of our system. We wanted to transfer as much of the math calculations onto the GPU. Coincidentally for loops were the main driving force processing our data which makes it suitable for parallelization. However a few operations requires EmguCV dll files for execution and this can only be done on the CPU. Additionally the for loops are for processing the captured color, depth and UV maps which are updated at every frame; the time it would take to transfer these images to the GPU for processing at every frame would be ,theoretically, too much time. This would also be limited by the GPU's fillrate and memory bandwidth. Overall this direction of work demanded too much time to get it where it was and needed more for it to be fully operational therefore we had to give it up in the end, and be content with the system's current performance. We weighed it out that with the remaining time it would be better to introduce more user friendly interactions with the system.

4.3.4 Increasing the camera's field of view by extending the camera distance

The Oculus Rift presents two images, one to each eye, generated by two virtual cameras separated by a short distance. The FOV of the virtual cameras must match the visible display area. If we use the term display field of view (dFOV), we are referring to the part of the user's physical visual field occupied by VR content. It is a physical characteristic of the hardware and optics. The other type of FOV is camera field of view (cFOV), which refers to the range of the virtual world that is seen by the rendering cameras at any given moment. All FOVs are defined by an angular measurement of vertical, horizontal, and/or diagonal dimensions. It is calculated/computed based on the eye distance and display size.

Right now with the DS325 camera mounted on top of the Oculus Rift DK2, its 320 by 240 aspect ratio UV map capture is only a small percentage of the DK2's 1920 by 1080 display resolution. As this is a hardware limitation of SoftKinetic's API for UV maps, the rendered image would be unclear and blurry if we were to stretch the image to cover the DK2's display resolution. But if we leave it as one-to-one with the real world, that is, not stretched in any direction, the image for real world visualization will be clear and sharp but still not covering the entirety of the DK2's display. A one-to-one with the real world means the perspective of the user looking at his hands in real life is the same as the perspective of the user looking at his hands through the Oculus Rift. One solution to this we thought might work is placing the camera behind but still above the user's head. The UV map resolution will still be 320 by 240, but now as it is placed behind, it is able to capture more of the real world surroundings. The camera's field of view is still the same but now with more surroundings in sight, we could stretch the image to give the illusion that they fit exactly one-to-one with the real world.

It is one thing to design it in theory and another to practically do it, which we tried by modifying the existing mount of the DS325 camera on the DK2. We tried extending the mount so that the camera is placed above and behind the DK2 about a metre, but there were a few problems: it made the Oculus Rift DK2 heavier which is already quite heavy itself, also the mount needs a better lock-in-place for its heavy weight to stay attached on to the Oculus Rift DK2. We had a rough first working prototype done, and were able to observe the changes in the UV map's captured surrounding content. It turns out

that we were only able to stretch the image a little for that cost of placing the camera one metre backwards. In the end, we decided to drop this idea, as we would need to extend the camera a lot longer and higher for most of the dFOV resolution to be covered, additionally it demands too much time for a good working mount to be produced.



FIGURE 4.10: Extended Mount First Prototype

4.3.5 UI for Implemented Gestures System

As mentioned in the section before this, we swapped the head-shaking feature to change the viewable depth distance for an easier type of interaction using hand gestures. Hand gestures are less dis-orientating especially when the viewer is already focused on the movie, vigorous head-shaking would mess up their concentration and tire the mind after a few attempts. Furthermore with the heavy weight of the Oculus Rift headset, shaking would be painful on the nose bridge. The gestures implemented were those already built into the SoftKinetic API, and they seem appropriate to what we plan to use them for, such as "Thumbs Up" would be to increase the depth distance and similarly with "Thumbs Down" to decrease. This will be explained more in section 4.3.2

Having so many points of interaction but no feedback would be very poor design as we would not know if the system has picked up our input or what it is doing. Among all types of feedback that can be provided, visual is the fastest and easiest which we solely focused on in our UI design. Since we have four types of gesture inputs, we need the same number of visual feedback to viewers. Also they should be easy to see, logical, and viewers should not feel that they are in their face.

Our first design was a text-based design as shown in 4.11, adapted from an existing SoftKinetic project sample. It looked fine but text are always hard to read, and requires some attention to read it which deviates the viewers from their movie. Thus we opted to use something more graphical like pictures, and the more familiar it is to viewers the more intuitive and easy it would be for viewers to recognise at first glance. We came up with two sets of images, one would be to indicate the system's detection of viewer's hand gestures; this would be the "BIG 5", "THUMB UP", and "THUMB DOWN" gestures; the other set are up and down arrows to show the system increasing (refer to figure 4.13), decreasing (refer to figure 4.15) or not changing the viewable depth distance (refer to figure 4.14). To create some kind of systematic organization of the UI elements, we arrange them in an order that made them look altogether as a heads-up display (HUD).

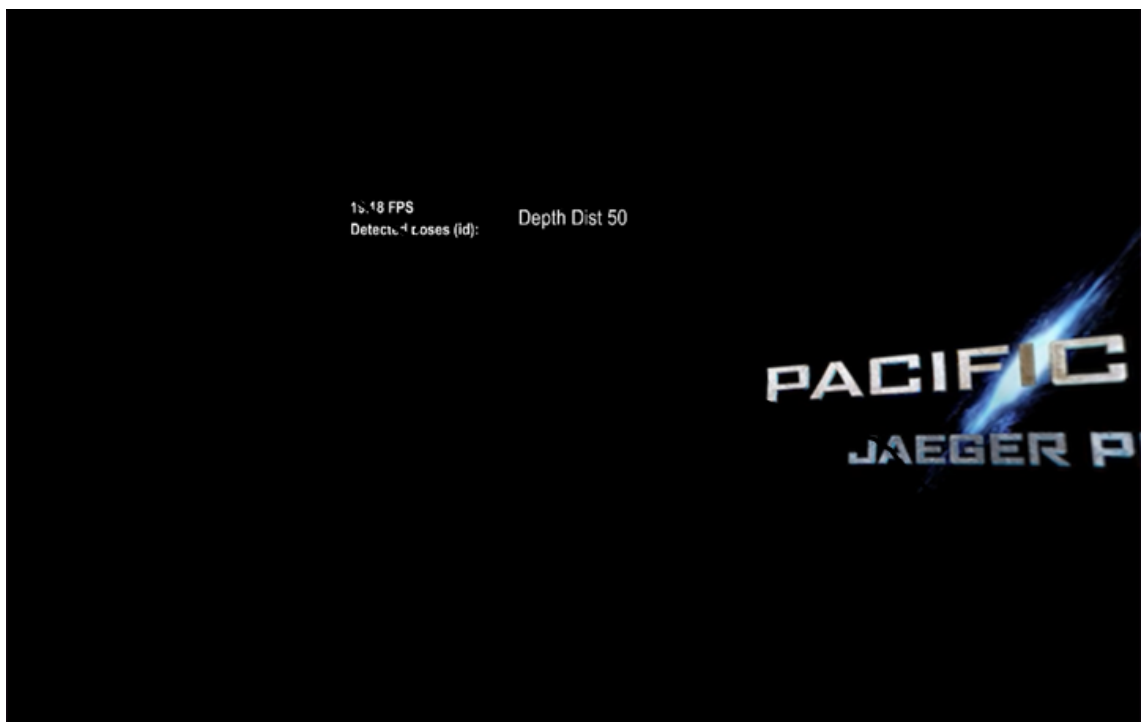


FIGURE 4.11: Text-based UI design⁶

⁶Legendary VR: Pacific Rim Jaeger Pilot <http://bit.ly/1XBskV0>

⁷Legendary VR: Pacific Rim Jaeger Pilot <http://bit.ly/1XBskV0>



FIGURE 4.12: Text-based UI design with detected gestures given as text feedback⁷



FIGURE 4.13: Graphics-based UI - Thumbs Up + Up Arrow⁸

FIGURE 4.14: Graphics-based UI - OpenHand⁹FIGURE 4.15: Graphics-based UI - Thumbs Down + Down Arrow¹⁰

For a more appealing effect, we have the corresponding gesture images fade in during first detection, and fade out after a while. The arrow images are just toggle shown and not shown because they are smaller compared to the gesture images and therefore the overall UI looks "cleaner".

⁸World Of Tanks: 1941 Battle 360 Reenactment <http://bit.ly/1KZ50Qa>

⁹The Hunger Games : VR Experience <http://bit.ly/1kIicwy>

¹⁰The Hunger Games : VR Experience <http://bit.ly/1kIicwy>

¹¹The Hunger Games : VR Experience <http://bit.ly/1kIicwy>



FIGURE 4.16: Graphics-based UI - Fading out Open Hand + Fade in Thumbs Down¹¹

4.3.6 Viewable distance reading at the bottom

VR experiences can lead to simulator sickness - a combination of symptoms clustered around eyestrain, disorientation, and nausea. Even with a flawless hardware implementation, improperly designed content can still lead to an uncomfortable experience.

Various physiological systems govern the start of simulator sickness: a person's overall sense of touch and position, or the somatosensory system; liquid-filled tubes in the ear, or the vestibular system; oculomotor system, or eye-movement muscles [35].

Anecdotal evidence has suggested simulator sickness is less intense when games contain fixed visual reference objects - such as a racecar's dashboard or an airplane's cockpit - located within the user's field of view. To match all game or movie scenarios, one could use a virtual nose or "nasum virtualis". It was found that participants tend not to notice its presence. Researchers used electro dermal activity (EDA) sensors to record electrical conduction across the skin, which is affected by sweating due to excitement, a proxy indication of simulator sickness. Measurements indicated EDA differences between subjects playing applications with the nose and without [35].

To design a good VR UI, one should put this into consideration which is what we did but we did not use a "nasum virtualis", instead we placed some noticeable text which participants refer to for the size of the viewable depth distance. You can see this in

figures 4.13, 4.14, 4.15, and 4.16. None of our participants, even those who had not use a head-mounted display before, pulled out due to dizziness or nauseousness. It was encouraged for them to voice out if they feel it in the slightest. In the end, everyone was pretty comfortable with it but we have no quantitative data to back this up.

Sometimes participants would want to not see the HUD anymore during some point into the movie. Therefore we added the "WAVE" gesture into the HUD UI for toggling the showing of this viewable depth distance value at the bottom centre of the viewer's field of view. The "WAVE" gesture is easily detected by the DS325 camera and is one of the most responsive gestures that participants themselves liked too.



FIGURE 4.17: Graphics-based UI - Waving + Show viewable depth distance¹²

4.4 Summary

Our desktop set-up consists of a desktop CPU with the Oculus Rift DK2 and the SoftKinetic DS325 mounted on top of it. The mount is designed and 3D printed. The Oculus positional tracker is not used, and the participants do not use the keyboard or mouse. Their only limitation is the length of the Oculus Rift HDMI display cable which restricts them from moving physically around too much. Sennheiser headphones are used to provide decent quality audio to match up with the movie. A much better stereo with

¹²The Hunger Games : VR Experience <http://bit.ly/1kIicwy>

¹³The Hunger Games : VR Experience <http://bit.ly/1kIicwy>



FIGURE 4.18: Graphics-based UI - Waving + Do Not Show viewable depth distance¹³

noise-cancelling type could have been used and immensely improved the VR experience but our budget was limited.

Because VR has been a fairly esoteric and specialized discipline, user testing is absolutely crucial for designing engaging and comfortable experiences. We were not sure how many user testers we would need to get the right system running with the tools and equipment we currently have. It was more of a trial and error situation where the system will be continually refined based on feedback from test participants till we were confident that the program is capable of answering our research questions.

Chapter 5

User Evaluation

5.1 Evaluation Goal

This chapter describes the work done from our first prototype and its progression to the next after levels of testing and evaluation were conducted. The goal of the experiment is described after which the experiment design explained. The information sheet and consent form used in the experiment are found in Appendix A along with the approval of the user experiment from human ethics, and all subjective questionnaires used, including during TEDx, can be found in Appendix B.

As sense of presence/immersion and enjoyment are key factors determining our success criteria, they are what we need to measure (and so are our dependent variables). Since they are based on feelings of users, direct feedback is all that is needed which can be in the form of questionnaires with e.g. Likert scale ratings. Our control/independent variables could be the visualization presented to the user in three forms:(without user body, with user body, with user body and environment) and also the type of interfaces which users interact with to control the visible depth threshold. Random variables like fatigue have to be limited such that it would not influence the results produced by the participants.

A simple survey with questions like the Likert scale rating can help measure the immersion felt by users. For example, "How easy was it to move between Reality and Virtual Reality ? (1 = very hard , 7= very easy)." [5] Something like the Steed-Usch-Slater questionnaire used for testing the sense of presence would work too [17].

If more than one 360 panorama movie is used to create more than one VR experience with different degrees of blending, there will be multiple conditions to take into consideration. After completing each set of conditions, participants could complete a questionnaire rating their overall satisfaction, immersion, level of distraction etc [36], followed with a think-aloud user feedback [16].

5.2 Focus Group

A focus group is a great quick way to understand and capture a rough picture of the general public's opinions. We carried one out here in the HIT Lab NZ with participants mostly from the lab, during the early stages of this work to see what people are looking for. Participants consisted of a mixture of backgrounds in programming, usability design, and psychology. The ability to change the movie outcome and scenes based on user interactions went well, with everyone in the group emphasizing the importance of a great story as the viewer is now "part" of it, and not just a spectator. As some people would have less out-going and adventurous personalities than others, there should be an option to be just a viewer and not a participant. They might be more comfortable with sitting back, relaxing and just watching, like a "traditional" normal cinema experience.

Another idea was for viewers to choose or even walk to their own perspective view of the movie at any time during the film. It would be a social experience if multiple viewers with different perspectives would be merged into one by the system at later scenes of the movie. This would encourage interactivity not only with the system but with viewers among themselves. It should be a collective experience, at no expense of anyone, such as if one choose not to participate and want to just sit back, they should be able to; likewise if they choose to, they should be able to interact and talk with friends beside them like the usual cinema experience or even better.

Many people felt that with the ability to see one's ownself, movies would be games now. It should not be limited to oneself only, but if people are able to see each other in the movie, it brings another level of interactivity where they could, for example, throw balls at each other. How much of the body should be shown should be a creative decision by the movie director in his/her making of the movie. The goal behind showing the

body should be towards immersion rather than realism; but viewers should still have the freedom to decide what parts of their body they want to see and show.

Also having good audio programming and mixing would bring the immersion to another level. 3D surround sound, audio point sources with ambient accompanying would definitely further enhance the sense of presence too.

A few examples of what the focus group thought would fit products of this system are: 1) Detective movies/games where you find clues and discuss with your friends, you would be like police officers working with your friends. 2) Relaxing movies or anything with spectacular/breathtaking views. 3) Adventurous movies like *Anaconda* where you will be helping the main character find the giant snake in the water.

5.3 TEDx Conference

TEDxChristchurch 2015: Think Again was a sold-out event held in the city of Christchurch early October with talented speakers and performers all over New Zealand coming to encourage wonder, uplift, and inspire combining possibilities with practicality. This year's event featured former deputy domestic policy adviser to President Clinton, Eric Liu; co-founder of Tesla Motors, Ian Wright; journalist Rod Oram; musician Jason Kerrison; comedian Michele A'Court; and right-to-die advocate Matt Vickers, highlighting the wide diversity of professions and backgrounds in one location. TEDxChristchurch is an independent event operated under license from TED.



FIGURE 5.1: TEDx Demo

Our TEDx conference demo was our first prototype system to be publicly shown. It consisted of all the same hardware mentioned before in previous sections. The blending/fading of the user body and real surroundings visualization was fully operational with the head-shaking feature as the controls over the blending. The demo consisted of a variety of movies from horror to action packed. We gave participants a choice of what type of movie they would like to watch so that non-horror fans would not coincidentally have a nightmarish experience. It was a success with many people experiencing something new and having fun. Unfortunately due to expected crowded area, we had to forgo the training in using the head-shaking interface. Therefore quite a few participants had difficulties getting the head-shakes detected by the system.

We were able to obtain some qualitative and quantitative feedback from 14 participants for our first round of analysis about the system. The questionnaire used was designed with a focus on capturing what the participants' thoughts and feelings are about each aspect of our system design. First we asked about what they thought about seeing their own body visualization appearing and disappearing into the movie scene, and how they felt about it. Attributes of 360 panoramic movies is important to explore into too; so we asked about content and the types of viewpoint. Lastly we asked about interfaces that they would like to use in changing their body and real surroundings visualization; and also how much control over this visualization they would like shared between the system.

In summary, most people like the idea of seeing your body blending in and out of the movie. One said "You feel like you are part of the action. The main problem I felt I didn't like about much is see my arms cut when the blended distance was too small.". These responses confirmed our doubts over the small field of view that the SoftKinetic DS325 camera could provide, affecting the immersion of our application. But many loved the idea, "It's totally awesome", "Love this technology and it's potential", "More interactive than a 2D environment. Greater detail. More information.", "because this could add so many things that you wanted into the real world. Really immersive video games". As one participant puts "I agree but needs to get better for gaming", we certainly do agree too and hopefully in the near future this idea can reach its full potential.

With so much hype and excitement with our system's technology. it was no surprise to find their ratings for engagement and fun pretty high. Refer to figure 5.2. Another of

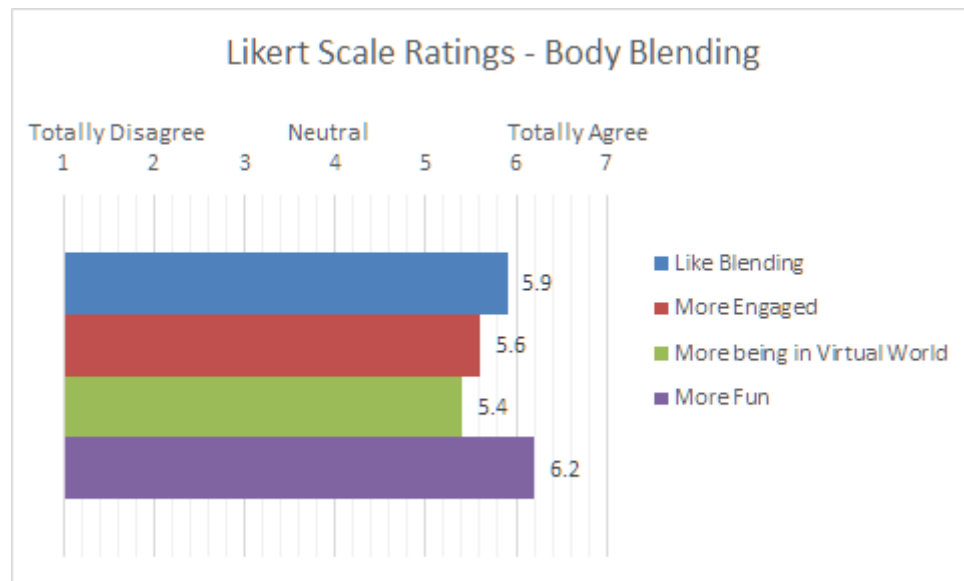


FIGURE 5.2: Participants Likert-scale Summary on Body Blending

our presumptions confirmed by results were that people would prefer to see their bodies blended in and out during live movies than animated ones. Refer to figure 5.3. From figure 5.5 observe that participants preferred the moving position of the movie camera to a fixed camera. I guess moving is more interesting and exciting for viewers while movie directors have to think carefully about their techniques as moving in a 360 movie with the viewer mostly stationary is likely to induce nausea. Regarding the timing and when should the movie show a viewer's body, many felt it should be relevant to the movie and purposely intended which altogether points back to the movie director. Thus it is important when shooting quality 360 movies to remember that viewers not only are watching but are "being" and "participating" in the movie.

The head-shaking feature was not well-liked by majority. Refer to figure 5.6. It barely pleased a few people who were able to get it working but with some effort. Most of the others tried head shaking too many times and ended up feeling too giddy to try again. Also shaking definitely makes it worse if they have tendencies to get nauseous. Many thought of using other types of interaction, hand gestures seems to be mentioned the most or doing something using hands. The main problem with using hands is you only have two, thus you can at most do two interactions at the same time, with one each hand. A participant suggested using voice commands, or even eyes blinking which adds another level of complexity to the system.

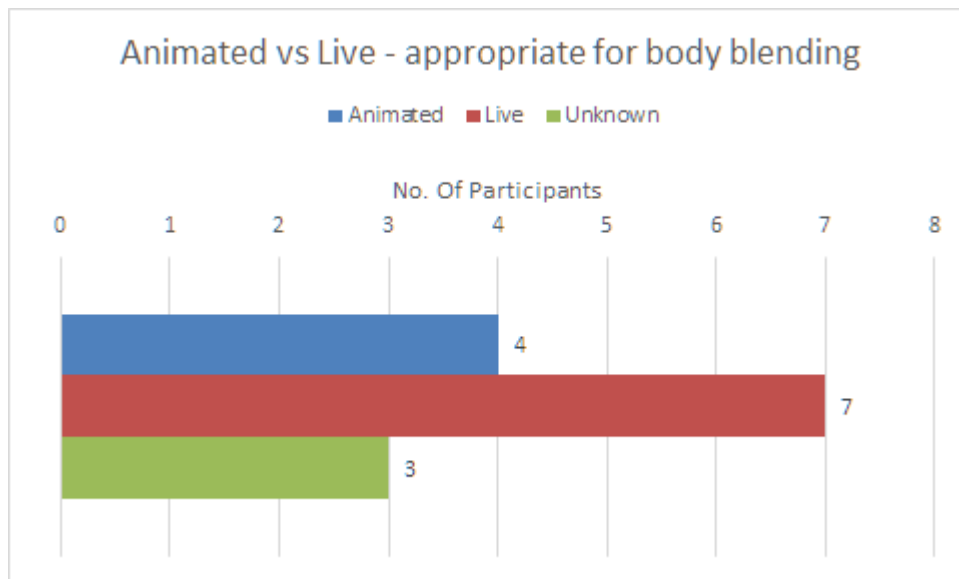


FIGURE 5.3: Participants Preference on Type of Movie

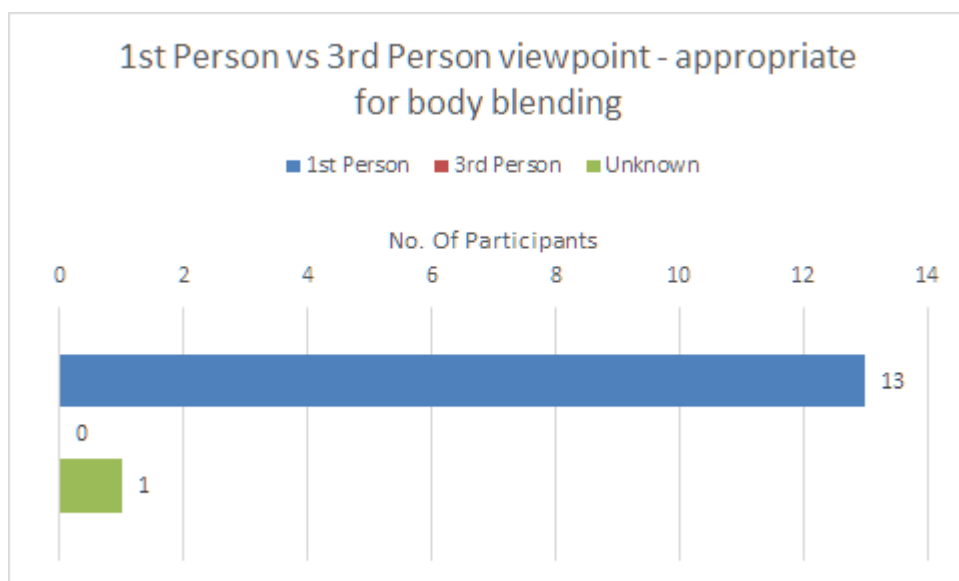


FIGURE 5.4: Participants Preference on 1st/3rd Viewpoints

Based on more participants' feedback we learnt that the genre of the movie does not really affect their likings, other than the common extreme dislikes of horror for many people. The genre of movie are too sparse and diverse to accurately point out which affects their likings especially when this is highly subjective. However the realism of the movie, that is, either it is animated or live does considerably affect participants' perception towards their level of engagement with the movie whenever they see their body appearing and disappearing into it which we can see in figure 5.3.

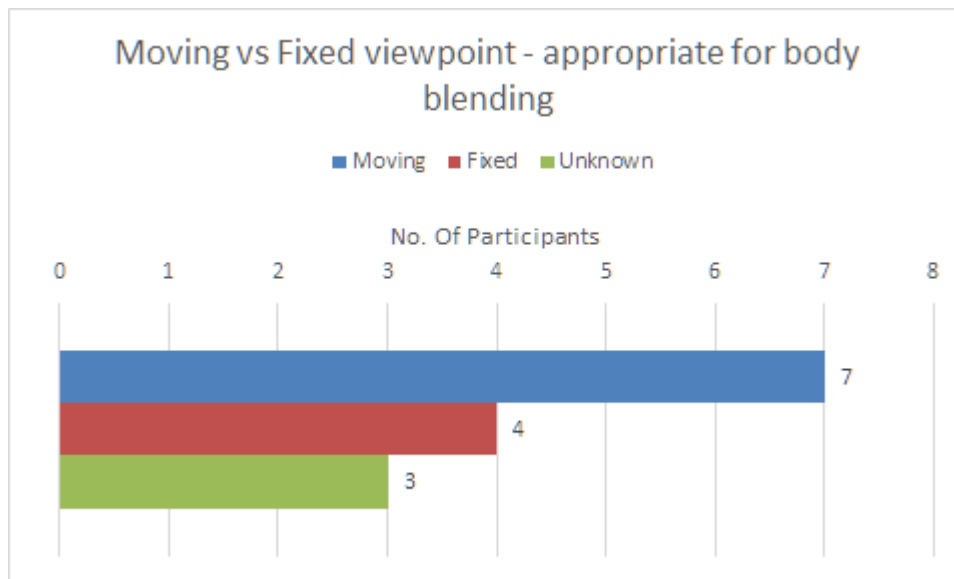


FIGURE 5.5: Participants Preference on Moving/Fixed Viewpoints

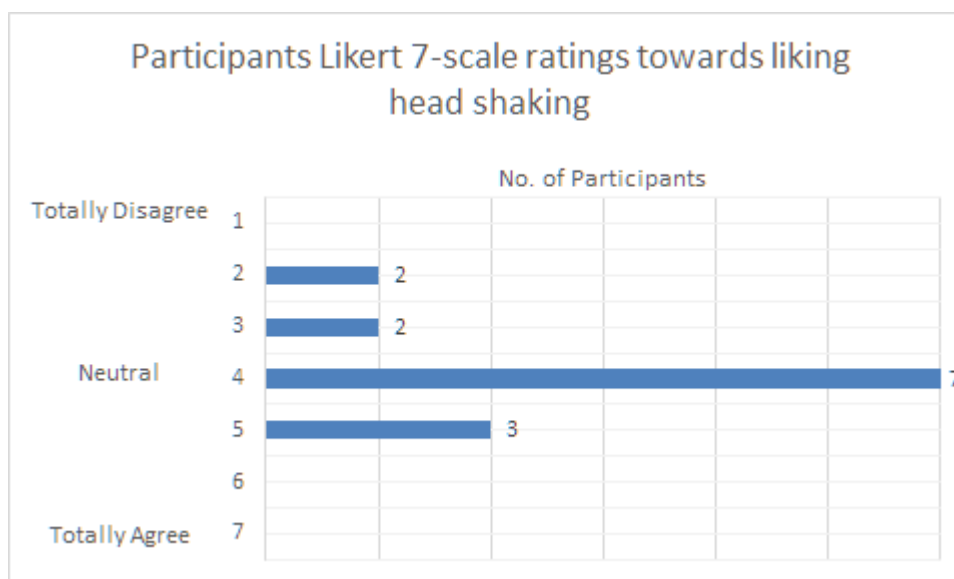


FIGURE 5.6: Participants Responses to Head Shaking

With the head-shaking interface implemented, the system then was providing a shared control between the user and system where when the user shakes his/her head they will be see more of the user body and surroundings visualization; over time this real world visualization will fade and disappear leaving the user in the virtual movie world if the user does not shake again. We asked participants if they liked that sort of control which they have experienced in the demo. We gave them five options to choose from: Manual control by the user only, Auto control by the system only, Current control which is the

head-shaking interface, Fixed control which means the user can see some visualization and the movie but they do not fade in or out, lastly we have other control which can be anything else than previously mentioned. Figure 5.7 shows their responses that they slightly preferred to have complete control with some liking the mutual participation with the system. One participant did not respond.

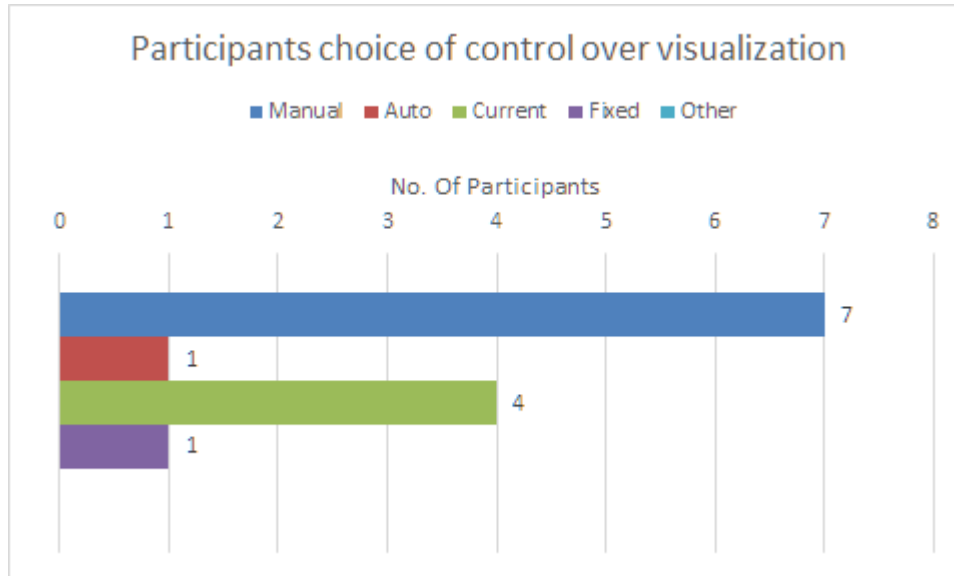


FIGURE 5.7: Participants Preference to Type of Control over Visualization

5.4 User Experiment Design

There are quite a few things we are curious about:

- if the appearing/blending of the body affects the participant's sense of presence
- if the realism of the movie affects the participant's sense of presence
- if the type of control over the appearing/blending of the body affects the participant's sense of presence

To test for the sense of presence, we use the IPQ tool designed and created by igroup¹. We also decided to add in engagement to our test measurements as it seems appropriate and might prove useful in our analysis [37].

¹igroup Presence Questionnaire IPQ http://www.igroup.org/pq/ipq/ipq_english.htm

Our whole experiment is split into two sessions, the first will consist of 4 conditions altogether coming from 2 levels of "Body present or not present", and 2 levels of "Live movie or animated movie". The first part's experiment design is a within-groups, 2 by 2 factorial design. Both factors are within-subjects factors since every participant will attempt all 4 conditions. There would be two main effects, and one interaction two-way effect to look out for. Being a within-group design, to minimise practice/learning effect, we have to counter-balance the order of presenting the conditions to every participant. The 4 by 4 balanced latin square design is incorporated here.

	A	B	C	D
Example,	B	D	A	C
	D	C	B	A
	C	A	D	B

where A = (Body present + Live action movie)
 B = (Body present + Animation movie)
 C = (Body not present + Live action movie)
 D = (Body not present + Animation movie)

The second part consists of 3 conditions from the 3 levels of "Type of control over blending/appearing of body". The experiment design is also a within-groups design. This only variable for part 2 is also a within-subjects factor since every participant will attempt all 3 conditions. Thus there is only one main effect to observe for. Similar to part 1 of this experiment, we need to apply counter-balancing to the order of conditions presented to participants. The 3 by 3 balanced latin square design is used.

	A	B	C
Example,	B	C	A
	C	A	B

where A = (Manual)
 B = (Automatic)
 C = (Shared)

5.4.1 Hypothesis

In the first session/part of the experiment, we have three hypotheses with each a null hypothesis and an alternative hypothesis. The first two are respectively regarding the each independent variable. The third is the interaction effect between these two IVs.

- First null hypothesis is: *There is no significant difference in sense of satisfaction/presence between seeing your body and not seeing your body blended together with the movie.*
- First alternative hypothesis is: *There is a significant difference in sense of satisfaction/presence between seeing your body and not seeing your body blended together with the movie.*
- Second null hypothesis is: *There is no significant difference in sense of satisfaction/presence between watching a live action movie and an animation movie.*
- Second alternative hypothesis would be: *There is a significant difference in sense of satisfaction/presence between watching a live action movie and an animation movie.*
- Third null hypothesis is: *There is no significant interaction between factors "Type-OfMovie" and "BodyShownOrNot".*
- Likewise from the previous two null/alternative hypotheses, the third alternative hypothesis is opposite of the null.

In the second session/part of this experiment, unlike the first we only have one hypothesis.

- The null hypothesis is: *There is no significant difference in sense of satisfaction/presence between manual, automatic, and shared control over the real world blending together with the movie.*
- Alternative hypothesis would be the opposite: *There is a significant difference in sense of satisfaction/presence between manual, automatic and shared control over real world blending together with the movie.*

5.4.2 Materials

In the first part of the experiment, we are mostly concerned that the movies presented are clearly separated between live action and animation. Live action movies are defined here as movies with real human actors/actresses in them; it is not limited by special effects or computer graphics but ultimately they should be based on the real world scenario such as *Lord Of The Rings*. Animated movies are completely computer generated films that can be cartoony or animation like *Despicable Me*. The genre of each movie might affect the participant's perception and liking towards it, therefore during the selection of our movies we narrowed down the spread of movie genres. At that time, 360 movie videos were gaining popularity among the internet community, with a massive number of our selected videos taken from YouTube. However the 360 movies were still at its infancy, and all sorts could be found ranging from inappropriate content to badly produced.

Another clear definition of the type of videos we were looking for was it has to be first-person perspective. From our TEDx qualitative feedback, most preferred a first-person view (13 out of 14) than a third-person (0 out of 14) as they felt more engaged and naturally more synced since they are "being" the main character in its position. As demonstrated in figure 5.8, we can see that the movie character is waving to the viewer. And it is this connection between the virtual characters and the viewer exchanging looks without saying anything that defines this type of 360 movie as a first-person. Edward Saatchi from Oculus Story Studio explain this best commenting "Is Henry doing that because I am here?" in their premiere video². Third-person, on the other hand, lets the viewer observe what is happening around the main character/s and there can be many variations of how the filming camera pans or moves. See figure 5.9 where the camera is up-close to the main character because most of the action is happening in a small space. While figure 5.10 most of the time has the camera far from the main character as it engages a large-sized giant character thus it uses more scene transitions than figure 5.9. Figure 5.11 shows a 360 movie also shot from a 3rd person perspective however it is purposely exploring cinematic conventions on view translations with questions: "Do camera cuts work in a 360 degree environment when the viewer is in control? Can the viewer's gaze be effectively directed? Is it possible to have the camera move without

²Oculus Story Studio : Henry <https://storystudio.oculus.com/en-us/henry/> <https://youtu.be/51ECD0-CV9o>

causing nausea?” These are interesting thoughts and relates to this research too for both 1st and 3rd person perspectives.



FIGURE 5.8: Movie with 1st Person Perspective³



FIGURE 5.9: Movie with 3rd Person Perspective⁴

³Oculus Story Studio : Henry <https://storystudio.oculus.com/en-us/henry/> <https://youtu.be/51ECD0-CV9o>

⁴Penrose Studios : The Rose & I <http://www.penrotestudios.com/> <http://kaleidovr.com/showcase/roseandi>

⁵Colosse VR <http://colossevr.tumblr.com/>

⁶Avram Dodson : The Last Mountain <http://kaleidovr.com/showcase/the-last-mountain> <http://ocul.us/1TxqpCw>



FIGURE 5.10: 2nd Movie with 3rd Person Perspective⁵



FIGURE 5.11: 3rd Movie with 3rd Person Perspective⁶

Also there is a common technique in the making of these movies where if the viewer looks down, he will notice he has a virtual body in the movie. However this virtual body is part of the movie thus it moves on its own which often confuses the viewer to whether the body is controlled by their own physical body such as in figure 5.12. Ideally it would be fantastic if that could happen, and is possible with our technology where we could dress the viewer's real body with virtual clothing. At where our development stands, it requires further work and development to reach that capability. Nevertheless this type of movie does not suit, we need a 360 movie that does not give the main character a virtual body so that we can give viewer's their real body. Refer to figure 5.13.



FIGURE 5.12: Movie with a Virtual Body⁷

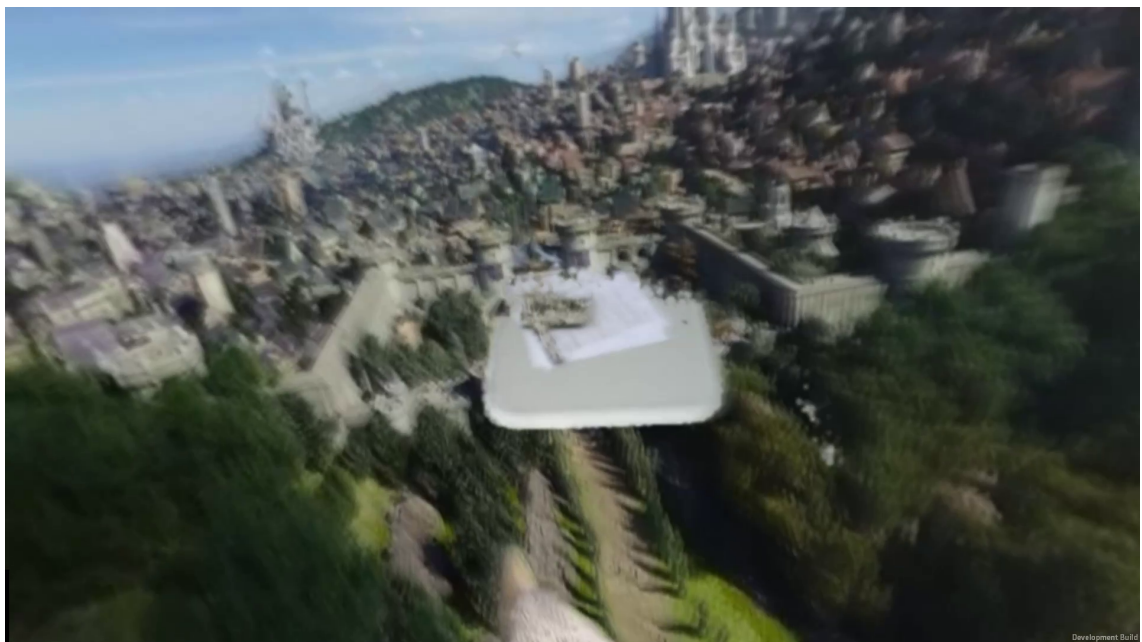
Another difference between the videos chosen to be used in the user experiment, are that one half has a stationary perspective and the other half, a moving perspective. Watching them normally might not make much of a difference but when incorporating user body and real surroundings visualization, it changes the whole experience dependent on which perspective the movie shows. One example of how things can get weird is when you are watching a movie with a moving perspective like a moving car and slowly you noticed your fixed real surroundings starts appearing. One example is in figure 5.14 could be the Warcraft VR trailer where the viewer would be riding on a flying eagle, it would be really weird that when you turn your head to suddenly see your real surroundings,

⁷Legendary VR: Pacific Rim Jaeger Pilot https://youtu.be/93X_65NS04o

⁸Clash of Clans 360 VR <https://youtu.be/wczdECcwRw0>

FIGURE 5.13: Movie with No Virtual Body⁸

such as a table, pop in beside you out from thin air. It would be different if your real surroundings en-covered your entire vision and "brought you back to reality", but if you see the fixed real chair beside you while you are in a virtual moving car, it does break the experience unless that was the movie director's intention.

FIGURE 5.14: Movie with Moving Perspective and real surroundings distraction⁹

⁸Legendary VR : Warcraft Skies of Azeroth <http://bit.ly/1HDH64c>

Since our first part/session of the experiment has 4 conditions for every participant to complete, that is 4 per-condition questionnaires to fill out. Not to mention, there is another post-questionnaire at the end of this session. With so much time spent in only the first part, we had to keep the total experiment duration in mind. In the 4 conditions, 2 conditions would have a live-action movie played, and the other 2 an animated movie. Instead of 4 different movies in total, we decided to give each participant only 2. This is to minimize the time spent in familiarising and training, this would be explained in section 5.4.3. Each movie would be at most 3 minutes long, and at least 1 and a half minutes. This is to avoid time consumption and tiring the participants; we want to ease them into the VR experience since most would be first-timers.

By the second part/session of the experiment, many would be used to the VR experience, and the probability of getting nauseous is less since they would have had a small break too. There are 3 conditions in session 2 however no questionnaires are needed to be filled out after every condition. Only a post-questionnaire for session 2 is given to participants at the end to fill out. Also a within-subjects factor here is the type of control over the blending/appearing of the user's body with the movie. For participants to fully experience and utilise the tool, they would need sufficient time to do so. Thus only one movie would be played to each participant, they could either be live-action or animated. And now they would be 5 minutes long at most, giving participants plenty of opportunities to use their control tool.

Overall we have two types of movies, each unique to the session's purpose. Each participant would experience two movies in session 1 and one movie in session 2. In total, we had sourced 8 different movies for session 1, and 4 movies for session 2. To randomise the order we give each participant their session 1 and 2 movies, we used an evenly distributed random real number generator. For session 1, there are 8 movies in total with 4 animated movies and 4 live movies. Since each participant has to watch an animated movie and a live movie, we can organize the 8 movies pairs, giving us 16 unique pairs of one animated and one live movie combinations. Next we order the 16 pairs using the random number generator, and give the first 16 participants their movies; we continue doing so similarly for the next participants. It is much simpler for session 2 where there are only 4 movies and each participant watch only a movie. Likewise, we order the 4 movies using the same random number generator for the first 4 participants, and continue on for the next 4 followed by more.

5.4.3 Procedures

1. We invite the participant to the lab study area where the information sheet and consent form is given to them to read and fill out. A pre-experiment questionnaire is later filled out by them too.
2. Once they have understood the writings, we give them a proper briefing of what is going to happen in the whole experiment. We also let them try out the Oculus Rift by putting them in an 360 VR still environment; mostly to give first-timers a first taste in wearing the headset, and experiencing what it is like. We encourage them to look around, and move if they would like to, such as turning around on the chair or getting up from the chair to look around and look at the chair; this is to give them the experience of looking at the real world through the eyes of the DS325 camera, and showing that when the real surroundings are presented to them through the Oculus, they could still touch the real surroundings reminding them that they are "still" in the real world.
3. They are then asked to preview the session 1 movies that they are about to watch, on the desktop monitor. This was to minimize or completely remove any novelty/surprise effect so that when they watch them again on the Oculus Rift, they would no longer rate their overall experience hugely based on the content of the movie.
4. Finally they will go through session 1 and perform the 4 conditions in the order already pre-defined for them. After experiencing each condition, they are asked to fill out a per-condition questionnaire to rate the experience. At the end of session 1, there is a post-questionnaire to sum up part 1 of the experiment which has some open-ended questions to get some text feedback from the participants.
5. In part 2 of the experiment, a video presenting the how-tos in performing the hand gestures is watched by the participants. This is the perfect time for them to know what they have and can do during session 2; it also is a time for them to ask questions if they do not understand what to do.
6. After watching the how-tos video, they are given a chance to practice the gestures they have watched and learnt about. We set them up with the Oculus Rift and DS325 camera, and put them through a demo experience which has plenty of time

for them to try out their gestures. Headphones are not used so that they could ask questions and the researcher could reply and guide them through any difficulties they might have.

7. Once they are used to the gestures controls, we put them through the 3 conditions in the order already pre-defined for them. No questionnaire is needed for them to complete after each condition. Only one post-questionnaire at the end of session 2 is given to sum it up and conclude the whole experiment with open-ended questions.
8. Before they complete this last questionnaire, the researcher would ask them a few rough questions regarding their experience and suggest improvements regarding the system technology and what could be improved. Any opinions and comments provided from the participant is encouraged to be included in this last questionnaire.

In session one, all participants sat down. And in session two, all participants stood up. They were welcomed to walk around and see their surroundings blend in and out or just stand still and look around. We made it based on their own freedom so as to not restrict them and their curiosity. Additionally enforcing too many rules might affect their sense of Presence too.

5.4.4 Measurements

Sense of Presence is the subjective sense of being in a virtual environment. Importantly, it can be differentiated based on the ability of a system's technology to immerse a user. While being based on technology, sense of presence is a variable of user experience. Therefore we used the Igroup Presence Questionnaire (IPQ), a scale for measuring the sense of presence experienced in a virtual environment (VE).

IPQ has three sub-scales and one additional general item. These three sub-scales emerged from principal component analyses and are independent of each other.

They are:

1. Spatial Presence: the sense of being physically present in the VE

2. Involvement: measuring the attention devoted to the VE and the involvement experienced
3. Experienced Realism: measuring the subjective experience of realism in the VE

The additional general item assesses the "sense of being there", and has high influences on all three factors especially on Spatial Presence.[38][39][40][41][42]

With hand gestures implemented in our system, we saw the chance of looking into more than just sense of presence; we wondered if interactivity could impact the user, specifically their engagement with the system. We wanted to see how well the system technology is able to engage users and provide them with an experience. Creating engaging user experiences is essential in the design of positive interactive system technologies. However the attributes of engagement are highly interwoven in, a complex interplay of user-system interaction variables. Thus we used a multidimensional instrument capable of measuring user engagement [37]. The survey scale comprised of six factors: Perceived Usability, Aesthetics, Novelty, Felt Involvement, Focused Attention and Endurability. These factors encompass the complex interaction between people and technology, comprising of disparate attributes (i.e.,attention, challenge, curiosity, intrinsic interest, and control)[43][44] curiosity, feedback, and challenge [45], aesthetics [46], and interaction, system format, and the presentation of the content [47].

However our questionnaire with all the IPQ questions was already pretty lengthy, therefore instead of all six factors we included only three, Focused Attention, Novelty, and Involvement. Take note that each factor has their own set of questions therefore we have to bear in mind of not overloading the existing questionnaire as we chose which engagement factors to include.

- Focused Attention would be important as it relates to users' perceptions of time passing and their degree of awareness about what was taking place outside of their interaction. It would be an interesting sub-scale to perform comparisons between with and without the body visualization as the participants interact with their real world surroundings which is outside of the virtual environment.
- Perceived Usability is not that crucial as it pertains to participants' emotions as they interact, also their perceived control and effort over the interaction.

- Aesthetics was the least important as we were not really concerned over the system's attractiveness and sensory appeal. Though an appealing user interface would be helpful in attracting people, which is why we designed visual gestures symbols and HUD, it was not a necessity in what we wanted to test for.
- Endurability seemed like not a good subscale in relevance to a movie experience as anyone would remember a good movie; also our supported interactions were few therefore cannot be considered hard to remember. Considering all these factors, we had to take note of the number of extra questions we are adding onto the questionnaire and thought we should not overload it.

The next two factors consisted of only six questions in total therefore were better candidates for including into the questionnaire than Endurability. They were Novelty and Felt Involvement.

- Novelty is the curiosity or participant's interest evoked; a movie experience can be a thinking experience causing viewers to be curious and think.
- Felt Involvement related to the participants' feelings of being drawn into and getting involved. An individual's salient needs and perceptions of fun are based on the level of importance, significance, or relevance [48] given to an experience by the individual.

5.4.5 Participants

Our participants were anyone who can watch a movie, comfortable with wearing a head-mounted display such as the Oculus Rift DK2, and have enough ability to communicate in English. We planned to recruit at least 24 participants and 48 the most. In the end we had 30 participants including pilot test subjects. We recruited participants through public advertisement such as posters pasted throughout the university, social media posts, and word of mouth. As we wanted to capture as much of the general public, preferably we hoped there would be a balance between both genders and a wide age range.

Within our 30 participants, we have 16 males and 14 females but our pilot test participants are both males, so our effective experiment participants consisted of 14 males and 14 females. The youngest would be 14 years of age, oldest is 32, and an average of about

22 years old. 18 out of 24 participants have never used a head-mounted display before. 5 only used it a few times and one use it weekly. Regarding the type of movie genres that participants prefer, science fiction and action-packed are the most selected, and the most hated is horror.

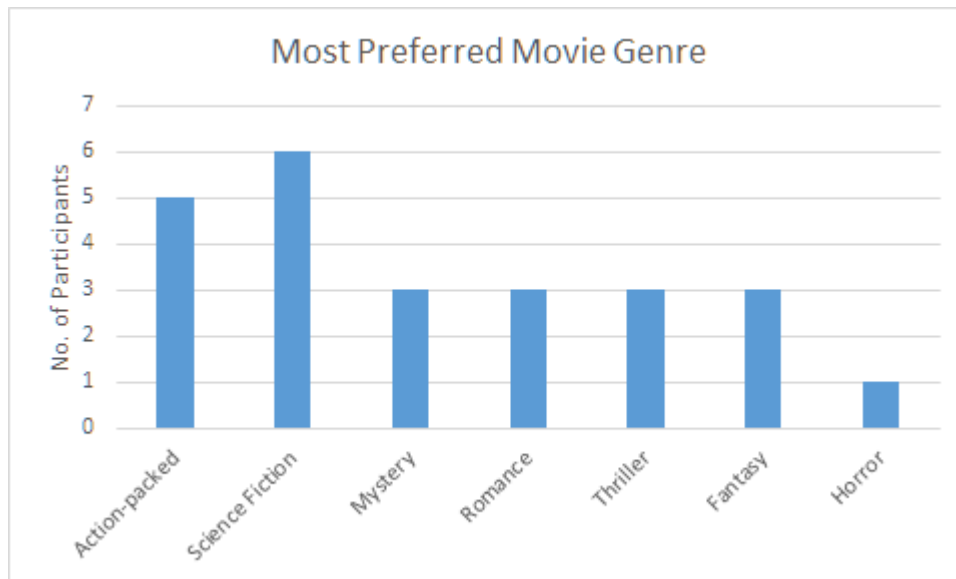


FIGURE 5.15: Participants Most Preferred Movie Genre

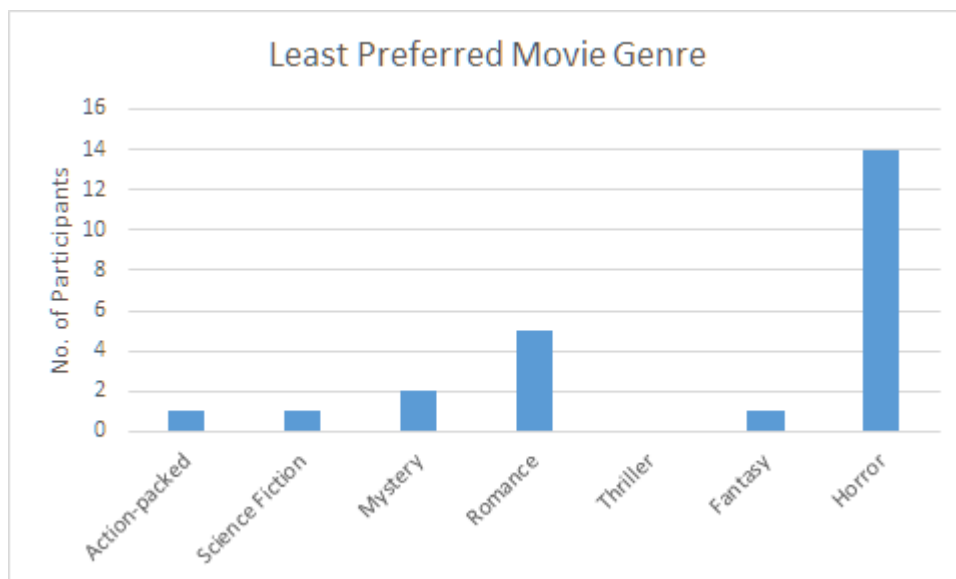


FIGURE 5.16: Participants Least Preferred Movie Genre

Chapter 6

Results and Discussion

6.1 Results

For every participant, there is a completed pre-experiment questionnaire, four per-condition questionnaires, one post-experiment-1 questionnaire, and a post-experiment-2 questionnaire. Most of our quantitative measurements were used in the per-condition questionnaires of session 1 of the experiment. Our qualitative measures were utilised in the two post-experiment 1 and 2 questionnaires. Results showed a significant difference in sense of presence and level of user engagement in the user body visualization factor though there were few who felt strongly that visualization alone was not enough and should consists of meaningful purposeful interactions. There were no significant differences discovered with the different types of movies, live or animated. Participants were undecided between manual and shared control with automated the least preferred.

6.1.1 Quantitative Measures

In session 1, we have two within-subjects factors, also known as independent variables, with two levels each, giving a total of 4 different conditions. A questionnaire containing IPQ and Engagement likert-scale questions are completed by participants after every condition. Since Likert-scale results are ordinal, we needed a non-parametric factorial analysis tool for repeated measures data. The Kruskal-Wallis and Friedman tests only handle one factor of N levels, and therefore cannot be used to examine interaction effects. Logistic regression can be used for some designs, but has assumptions about the

independence of responses and is often not suitable for repeated measures. The popular Rank Transform (RT) method of Conover and Iman [49] applies ranks, averaged in the case of ties, over the entire data set, and then uses a parametric ANOVA on ranks, resulting in a nonparametric factorial procedure. However, it was determined that this process only produces accurate results for main effects, and interactions are subject to big increases in Type I errors—claiming statistical significance when there is none [50] [51]. Therefore the Aligned Rank Transform (ART) procedure was devised and is what we used to transform our raw likert scale results before we could apply factorial parametric two-way repeated measures ANOVA in SPSS. We analysed the participants’ IPQ and Engagement values separately. Therefore for each, there are two main effects and one interaction effect resulting from studying the two within-subject factors, “Type of Movie” and “Body Shown Or Not”.

6.1.1.1 IPQ

Initially with 28 participants, we had no significant results: the “Body Shown Or Not” main effect ($F(1,27) = 2.506$, $p = 0.125$), and the “Animated or Live” main effect ($F(1,27) = 1.31$, $p = 0.262$) and interaction effect between these two factors ($F(1,27) = 0.157$, $p = 0.695$). However we noticed that most participants rated their “Body Shown” experience slightly greater than “Body Not Shown”; and there were a few who rated their “Body Shown” experience a lot lesser than “Body Not Shown”. Qualitative feedback though showed everyone was in favour of the “Body Shown” experience; if meaningful purposeful interactions with the body and surroundings visualization are implemented, this would be more.

	IPQ					S1 Conditions	
	1	2	3	4		1	Animated + Body not present
P18	5.428571	4	6	5		2	Animated + Body present
P22	5.142857	3.714286	4.714286	4.214286		3	Live + Body not present
P28	5.571429	4.142857	5	3.571429		4	Live + Body present
P30	6.5	4.285714	5.714286	5.357143			

FIGURE 6.1: Excluded Participants IPQ ratings

Excluding 4 participants and considering only 24 for analysis, with ART two-way repeated measures ANOVA ($\alpha = 0.05$), we found no statistically significant difference among the two “Type of Movie”, Live and Animated in terms of sense of presence ($F(1,23) = 1.274$, $p = 0.271$). We found statistically significant difference among the

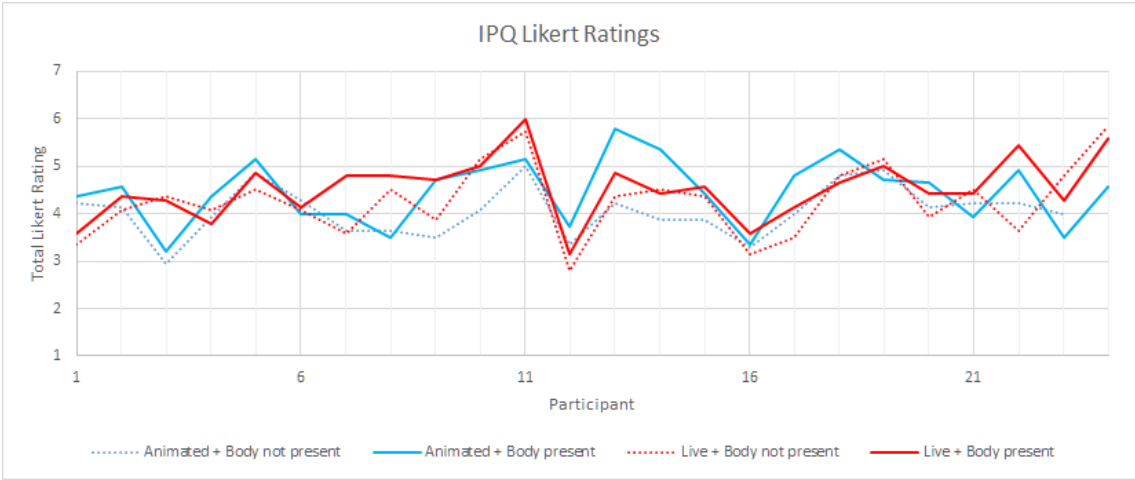


FIGURE 6.2: Participants IPQ ratings

S1 Conditions	
1	Animated + Body not present
2	Animated + Body present
3	Live + Body not present
4	Live + Body present

FIGURE 6.3: Session 1 Conditions Legend

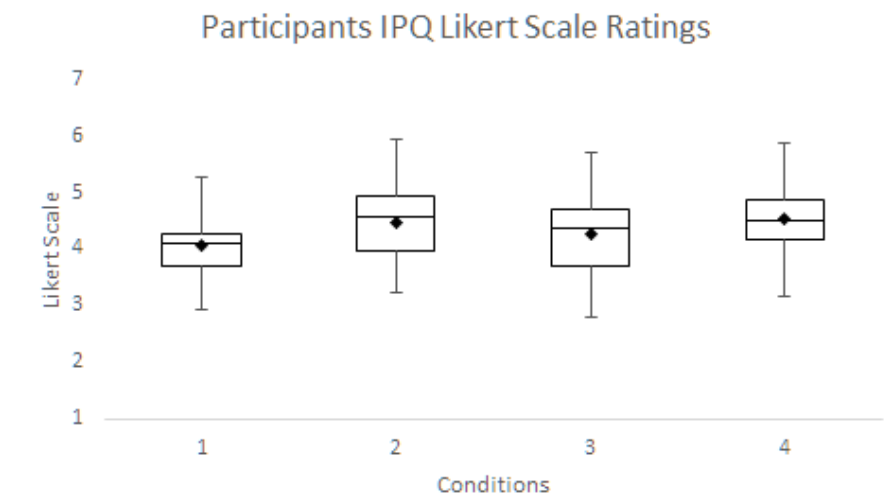


FIGURE 6.4: Participants Session 1 IPQ Box-plot Likert Ratings

	F	Sig.
Type Of Movie	1.274	0.271
Body Shown Or Not	12.563	0.002
Type Of Movie * Body Shown Or Not	1.775	0.196

FIGURE 6.5: IPQ Statistics Results ART 2-way Repeated Measures ANOVA

“Body Shown Or Not” in terms of sense of presence ($F(1,23) = 12.563, p = 0.002$). However we found no statistically significant interaction effect between “Type Of Movie” and “Body Shown Or Not” in terms of sense of presence ($F(1,23) = 1.775, p = 0.196$). Refer to figure 6.5.

6.1.1.2 Engagement

Likewise, we observed no significant difference among all two main effects and interaction effect: the “Body Shown Or Not” main effect ($F(1,27) = 1.324, p = 0.26$), and the “Animated or Live” main effect ($F(1,27) = 0.307, p = 0.584$) and interaction effect between these two factors ($F(1,27) = 2.286, p = 0.898$). However there was a difference when the same few participants 6.6 were excluded to reduce 28 participants to 24.

	Engagement					S1 Conditions	
	1	2	3	4		1	2
P18	4.769231	3.692308	4.769231	4.384615		1	Animated + Body not present
P22	4.153846	3.538462	3.923077	3.769231		2	Animated + Body present
P28	4.307692	3	3.923077	2.846154		3	Live + Body not present
P30	4.846154	4.230769	4.692308	4.230769		4	Live + Body present

FIGURE 6.6: Excluded Participants Engagement ratings

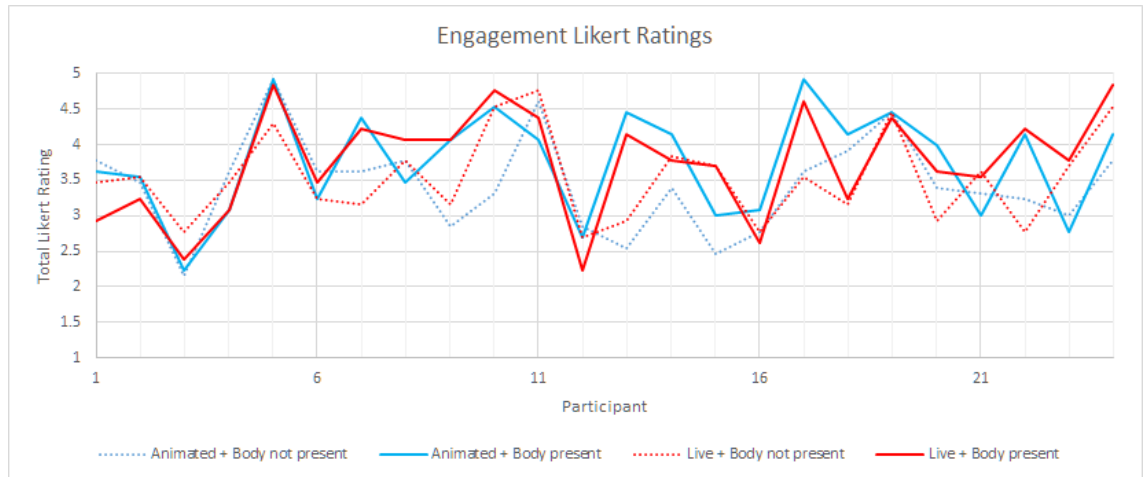


FIGURE 6.7: Participants Engagement ratings

Considering only the 24 participants, with two-way repeated measures ANOVA ($\alpha = 0.05$), we found no statistically significant difference among the two “Type of Movie”, Live and Animated in terms of sense of presence ($F(1,23) = 0.157, p = 0.696$). We found statistically significant difference among the “Body Shown Or Not” in terms of sense of presence ($F(1,23) = 5.183, p = 0.032$). However we found no statistically significant

S1 Conditions	
1	Animated + Body not present
2	Animated + Body present
3	Live + Body not present
4	Live + Body present

FIGURE 6.8: Session 1 Conditions Legend

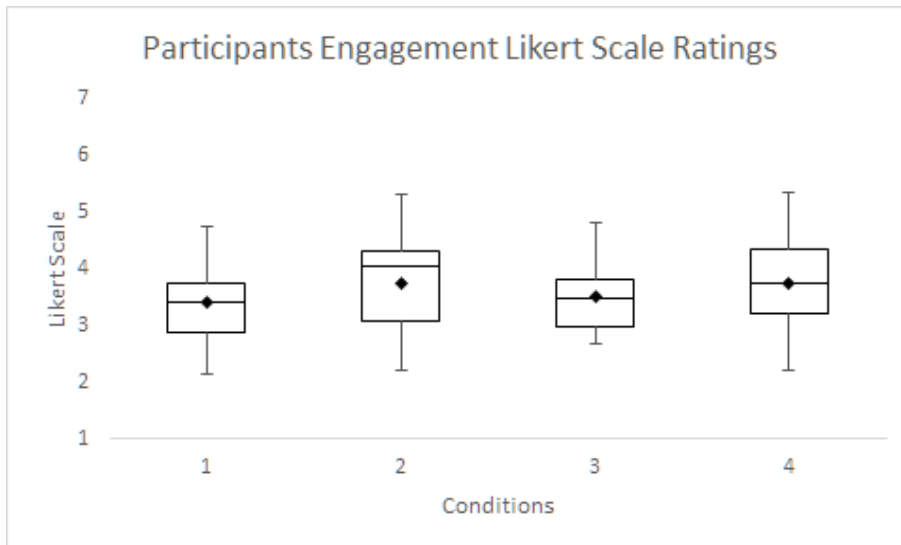


FIGURE 6.9: Participants Session 1 Engagement Box-plot Likert Ratings

	F	Sig.
Type Of Movie	0.157	0.696
Body Shown Or Not	5.183	0.032
Type Of Movie *		
Body Shown Or Not	0.394	0.536

FIGURE 6.10: Statistics Results ART 2-way Repeated Measures ANOVA

difference in the interaction effect between “Type Of Movie” and “Body Shown Or Not” in terms of sense of presence ($F(1,23) = 0.394$, $p = 0.536$). Refer to figure 6.10.

6.1.1.3 Session 1 Post-experiment

We also obtained some quantitative feedback regarding the whole of session 1 after completion. This is what we learnt:

Here we asked “Please choose which type of movie was more preferable to see your real body appear into: Animation Film or Live Film?”, and participants answer by rating from 1 to 7 where 1 is “Live Film”, 4 “Both” and 7 “Animated film”. From the results,

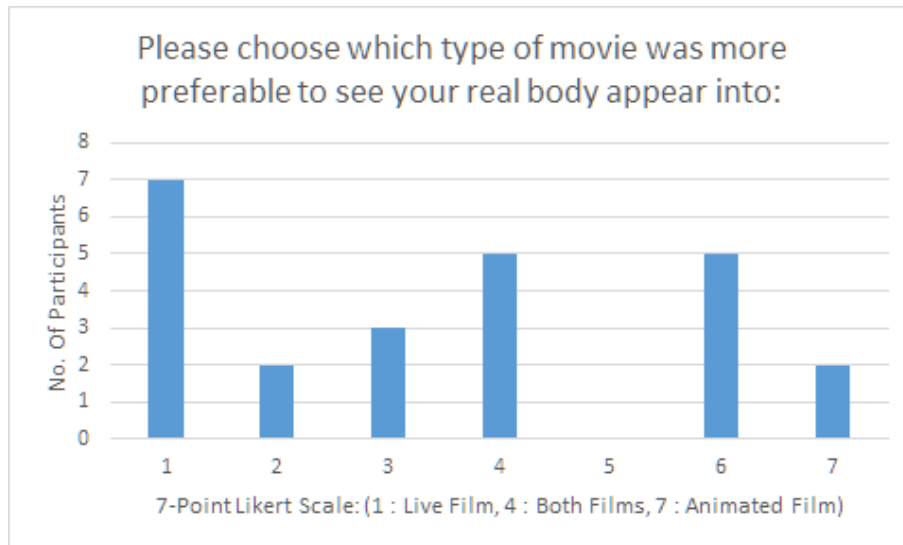


FIGURE 6.11: Participants responses to type of movie appropriate for user body visualization

Descriptive Statistics								
	N	Mean	Std. Deviation	Minimum	Maximum	Percentiles		
						25th	50th (Median)	75th
Average	24	4.0000	0.00000	4.00	4.00	4.0000	4.0000	4.0000
ParticipantsValues	24	3.5000	2.14679	1.00	7.00	1.0000	3.5000	6.0000
Ranks								
	N	Mean Rank	Sum of Ranks	a. ParticipantsValues < Average				
ParticipantsValues - Average	Negative Ranks	12 ^a	10.42	125.00	b. ParticipantsValues > Average			
	Positive Ranks	7 ^b	9.29	65.00	c. ParticipantsValues = Average			
	Ties	5 ^c						
	Total	24						
Test Statistics ^a								
	ParticipantsValues - Average	a. Wilcoxon Signed Ranks Test						
Z	-1.230 ^b	b. Based on positive						
Asymp. Sig. (2-tailed)	.219							

FIGURE 6.12: Statistics Results Wilcoxon Signed Rank

we see here that the average (3.5) tends to lean towards “Live Film” and “Both” with “Live Film” being chosen the most (7 responses) compared to “Both” (5 responses) and “Animation Film” (2 responses). However a Wilcoxon Signed Rank test showed that there is no significant difference ($Z = -1.23$, $p = 0.219$) between the participants’ likert scale ratings and the neutral likert value (4) as demonstrated in Figure 6.12. These are some of the participants’ comments following up from this question.

1. “because in the live film, I am able to see my body as it is now. It feels a lot

more real in the live film because my body appears in the film as a live body. In the animated film, it does not feel right to have a live body in a animated environment”

2. “My real body matched the live film more than the animation film”
3. “Live film make me feel more real than the animation film. The scene in the animation film looks very different from my body.”
4. “I didn’t fit in the animation film as well as in the live film. Seeing my real body in the animation films just reminds me that it’s just a virtual world.”

The comments above depict how the visual looks is important to the participants and that a variation in graphics visualization can appear more realistic in different types of movies. As the real body visualization is captured like a “live-feed” of the participant, it is part of the real world therefore observed as more suitable in a live type movie. Though the following comment from a participant, shows that some people found that there were no obvious difference especially when the live type film and animated film looks kind of similar since sometimes animation characters can be modelled exactly as humans. Similarly, we could modify viewer’s own real body visualization with virtual clothing and accessories to match the theme of the movie no matter if live or animated.

5. “Felt no real difference between the two in terms of realism”
6. “I feel like with animation films they are more interesting and people can be more creative with the whole animation idea. Whereas, I personally find it harder to engage with a live film.”
7. “If my real body was customised as one of the animated characters, it feels as if I’m more involved in the virtual world.”

Overall the visualization detail and quality of the user body, surroundings and the graphics of the movie, all influence the viewer’s perception towards feeling immersed, “being there” and “part of that world”.

Here we asked a similar question but now regarding with the real surroundings instead of the user’s body. “Please choose which type of movie was more preferable to see your real surroundings appear into: Animation Film or Live Film?”, and participants

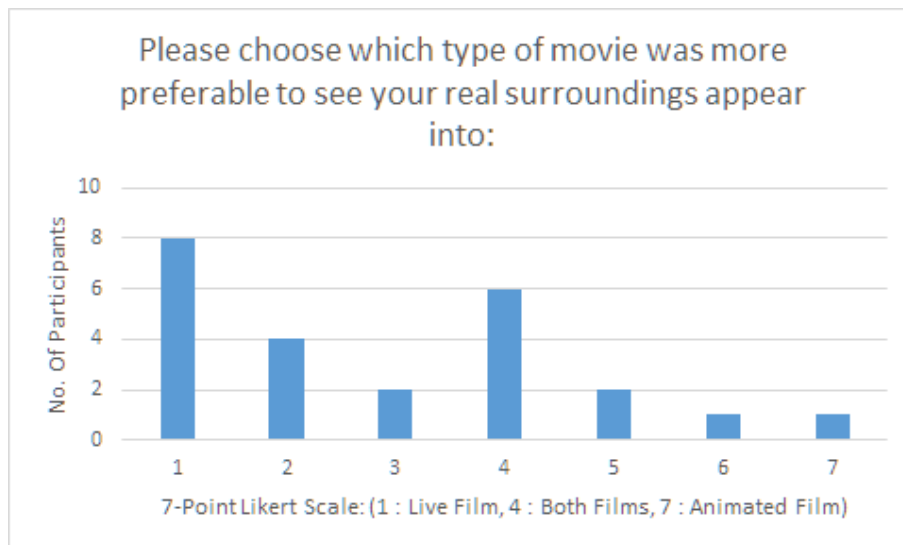


FIGURE 6.13: Participants responses to type of movie appropriate for surroundings visualization

Descriptive Statistics								
	N	Mean	Std. Deviation	Minimum	Maximum	Percentiles		
						25th	50th (Median)	75th
Average	24	4.0000	0.00000	4.00	4.00	4.0000	4.0000	4.0000
Participants\Values	24	2.8750	1.80127	1.00	7.00	1.0000	2.5000	4.0000

Wilcoxon Signed Ranks Test								
Ranks								
		N	Mean Rank	Sum of Ranks				
Participants\Values - Average	Negative Ranks	14 ^a	10.36	145.00				
	Positive Ranks	4 ^b	6.50	26.00				
	Ties	6 ^c						
	Total	24						

a. Participants\Values < Average

b. Participants\Values > Average

c. Participants\Values = Average

Test Statistics ^a								
		Participants\Values - Average						
Z		-2.639 ^b						
Asymp. Sig. (2-tailed)		.008						

a. Wilcoxon Signed Ranks Test

b. Based on positive ranks.

FIGURE 6.14: Statistics Results Wilcoxon Signed Rank

answer by rating from 1 to 7 where 1 is “Live Film”, 4 “Both” and 7 “Animated film”. Likewise from seeing your real body, seeing your real surroundings gave an average (2.88) leaning towards “Live Film” and “Both” with “Live Film” being chosen the most (8 responses) compared to “Both” (6 responses) and “Animation Film” (1 response). A Wilcoxon Signed Rank test showed that there is a significant difference ($Z = -2.639$, $p = 0.008$) between the participants’ likert scale ratings and the neutral likert value (4) as demonstrated in Figure 6.14. These are some of the participants’ comments following up from this question.

1. “they both did felt okay to have the real surroundings into the films, but it seems more normal to have real surroundings appear in the live film”
2. “The scene in the animation film looks very different from the real surrounding.”
3. “Real items in the real world would fit better in the live action film however it has to keep with the theme of the film otherwise it would not be realistic.”
4. “I think having a chair in the virtual world (from the real world) will interrupt with the animated movie. having surroundings into the live movie is more realistic.”

Just like above about the user body, likewise the participants’ experience with real surroundings visualization is dependent on the the visuals. A live film is much more relatable than an animated film if real surroundings were to appear. However nothing beats an engaging and interesting experience through the content and features offered by the movie, as told in the next few participants’ comments.

5. “I didn’t really experience much of my surroundings in the movies, I was just concerned about the stuff in the movies more than my surroundings”
6. “In the animation film we’re stationary, which is more intuitive this way vs the live film on which I’m flying on a couch.”

Most of the feedback here reflects the motivation for future work. For example, the chair taken from the real world surroundings could be used as the saddle on top of an giant eagle in the animated movie. The system could work such that once the camera recognises the chair which has been pre-calibrated and inserted into the recognition database, the system could decorate the chair with virtual clothing and materials causing the chair to look like an eagle saddle.

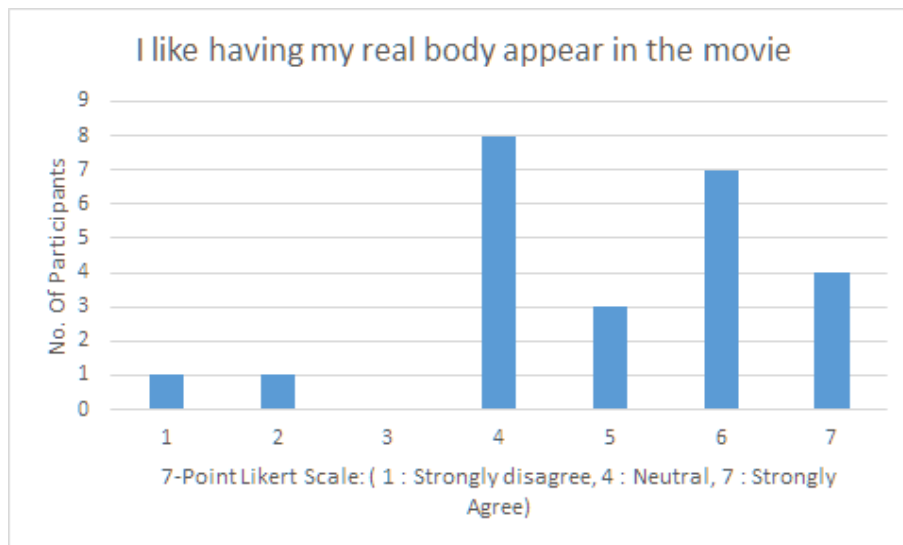


FIGURE 6.15: Participants responses to seeing their body appear in the movie

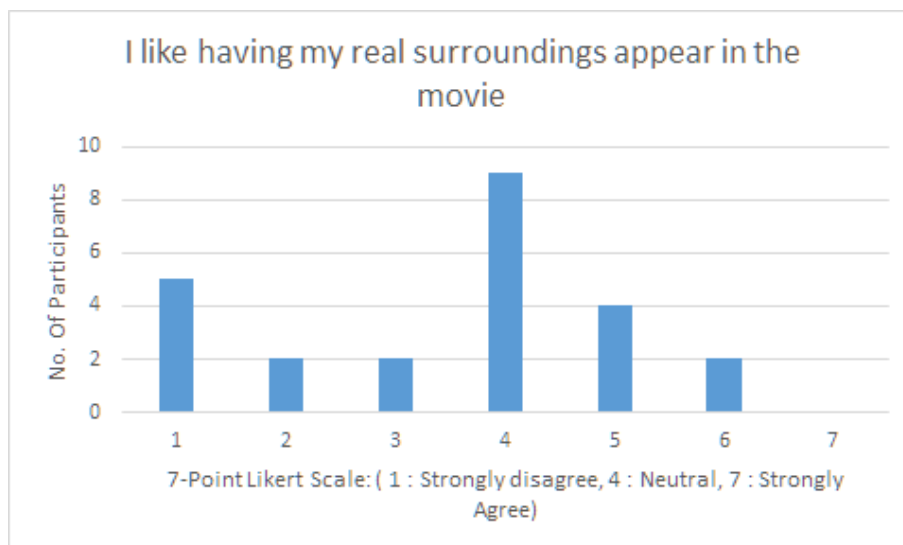


FIGURE 6.16: Participants responses to seeing their surroundings appear in the movie

Figures 6.15 and 6.16 illustrates the responses from all 24 participants regarding their thoughts on seeing their real body and surroundings visualization in the movie. Based on the histogram graphs we could see that the participants were more keen on seeing their real user body visualization appear in a movie, than their real physical surroundings. Figures 6.17 and 6.18 looks into whether participants get frustrated not controlling when and how much of the real body and surroundings visualization can be seen. Huge majority were neutral about wanting to see their body when they could not and/or not wanting to see their body when it is already shown, though there were more occurrences

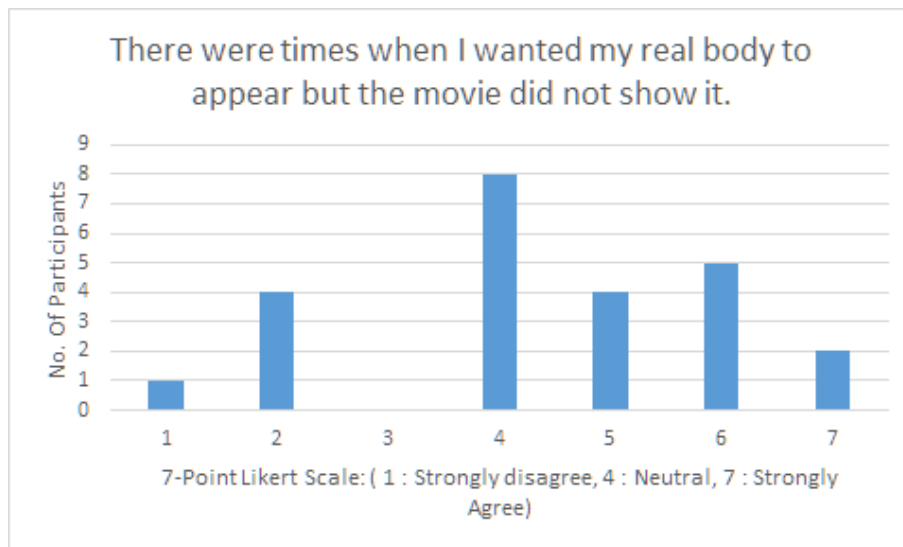


FIGURE 6.17: Participants responses to seeing their body when they did not want to

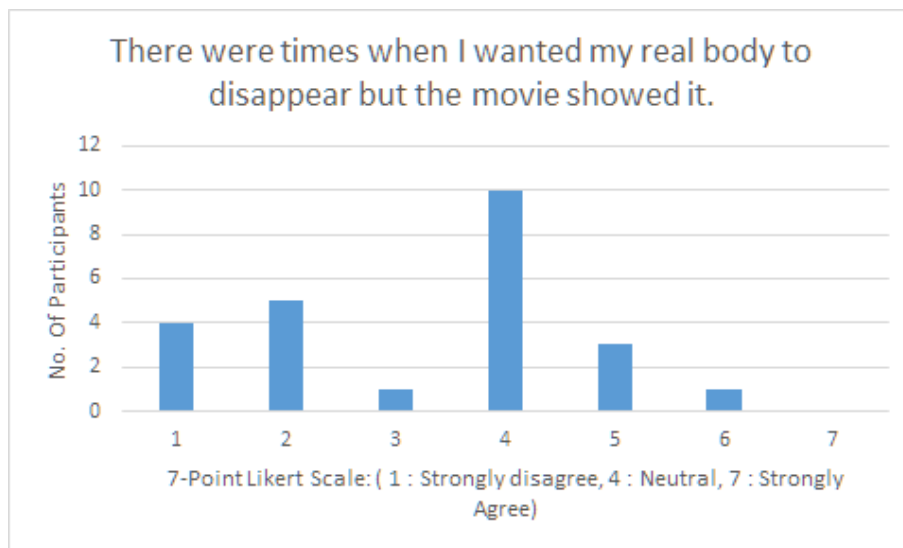


FIGURE 6.18: Participants responses to not seeing their body when they want to

when they wanted to make their body appear than disappear. We finally ask them directly regarding controls over their body and surroundings visualization and their answers are as shown in figures 6.19 and 6.20. As shown, participants were more towards having the control all to themselves and a majority were undivided on whether the movie should have control too.

To sum up all questions regarding the control over their real body visualization, we asked “What type of control would you like to have over your real body appearing in the movie?”, and participants responded by either “Manual control”, “Automated control”,

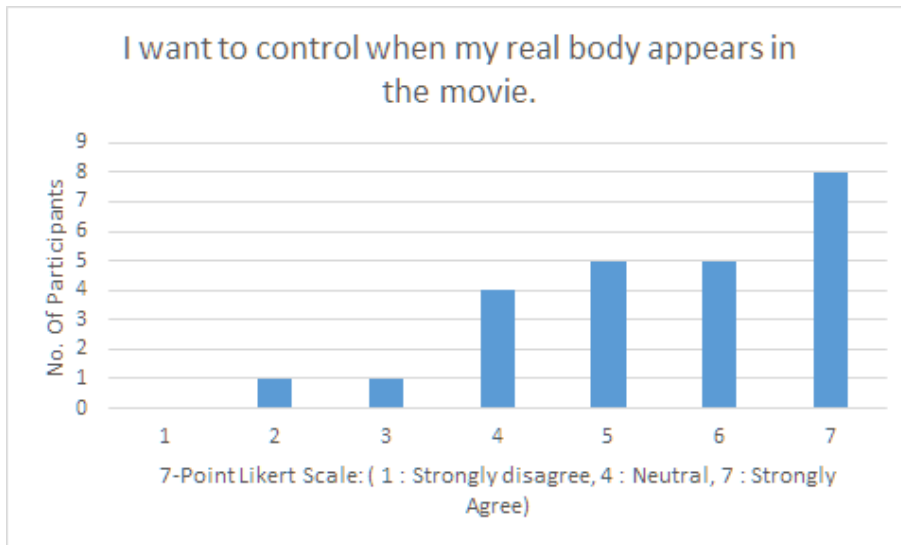


FIGURE 6.19: Participants responses to having manual control over body visualization

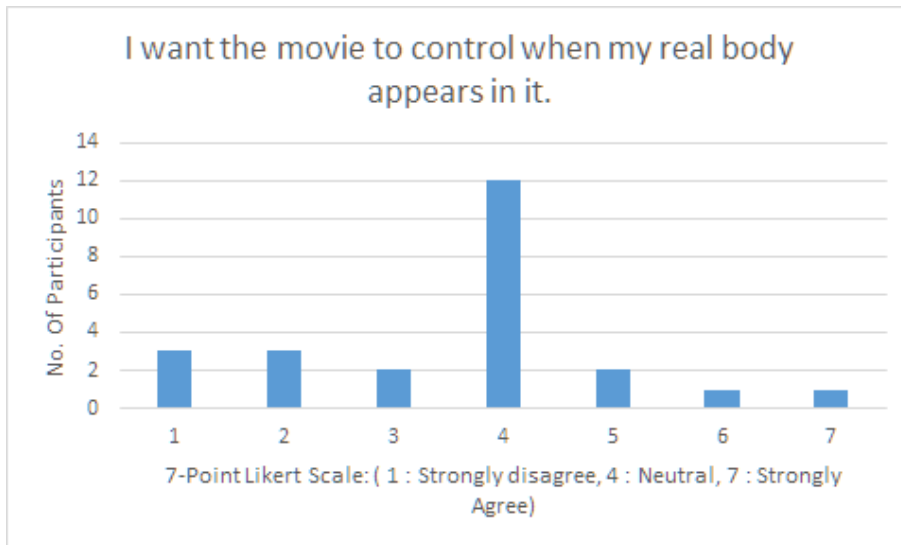


FIGURE 6.20: Participants responses to the system having automated control over body visualization

“Shared control” and “Fixed control”. Participants answered with “Shared control” being chosen the most (14 responses) compared to “Manual control” (9 responses), “Automated control” (1 response), and “Fixed control” (0 responses). A Chi-square Goodness of Fit test showed that there is a significant difference ($\chi^2(3) = 22.333, p = 0.000$) in the preference of the type of control as demonstrated in Figures 6.21 and 6.22.

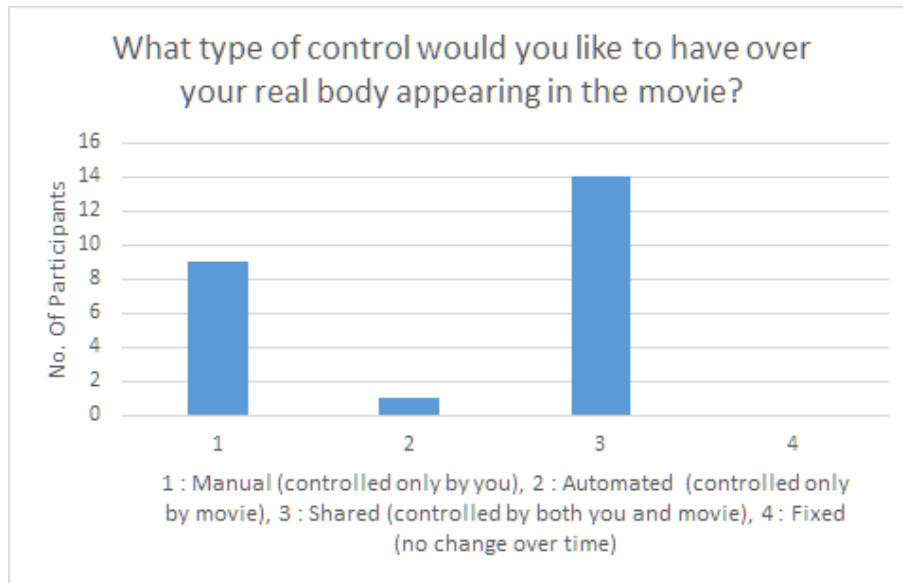


FIGURE 6.21: Participants responses to which type of control they want

	Control_type			
	Category	Observed N	Expected N	Residual
1	Manual	9	6.0	3.0
2	Automated	1	6.0	-5.0
3	Shared	14	6.0	8.0
4	Fixed	0	6.0	-6.0
Total		24		
Test Statistics				
	Control_type			
Chi-Square	22.333 ^a			
df	3			
Asymp. Sig.	.000			

FIGURE 6.22: Statistics Chi-square Goodness of Fit test on type of controls over visualization

6.1.1.4 Session 2 Post-experiment

We also obtain some quantitative feedback regarding the whole of session 2 after completion. This is what we learnt:

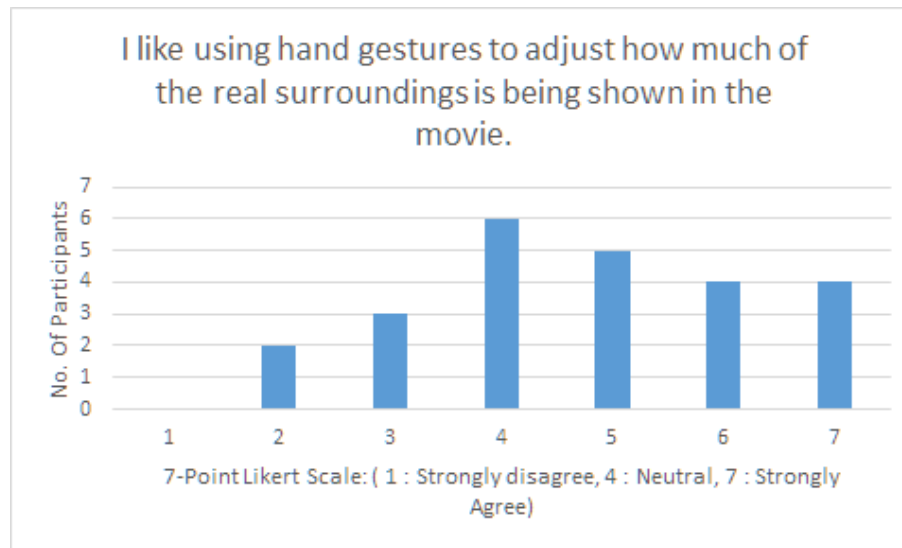


FIGURE 6.23: Qn2 Participants responses

We can see here in figure 6.23 that the average (4.75) tends to lean towards "Neutral" and "Strongly Agree". These are some of the participants' comments regarding what other interaction methods they thought would be better instead of hand gestures:

1. "some way to instantly toggle the appearance of the real world in the movie. Buttons to press in case the camera fails to detect the hand gestures"
2. "I guess it will depend on the movie. If the movie is action-packed, probably better to use something else as a gesture like a voice command/keyword. Rather than, hand gestures. But for other movies it is probably okay to use hand-gestures."
3. "buttons of some kind (e.g something like a volume button style), voice control"

Buttons and voice commands with keywords are a common alternative suggestion to hand gestures. Some thought the hand gestures are pretty good for certain limited-interactivity films, but if the movie were to be like a game, more responsiveness and controllability are needed thus the traditional keyboard would be best in this case. Another suggested using the swiping hand gesture but instead of toggling the appearance of the HUD, it would "swipe" through set-ups such as "now

see only virtual world”, “now see only real body visualization in virtual world”, and “now see real body and surroundings, that is, only real world”.

4. “I think hand gestures are the best for watching/being in a film. But if this was to be for gaming then using a keyboard would be really good.”
5. “I think the increasing and decreasing the threshold incrementally was clunky and unlikely to be helpful. Swiping between modes could be a better interface. Interacting with a game would also be cool, so that you changed how the content moved around you.”

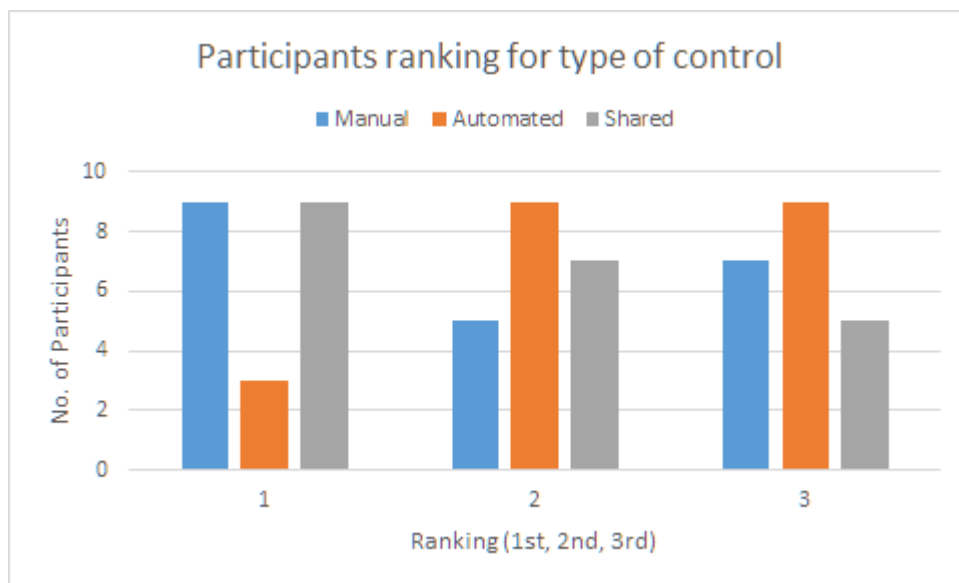


FIGURE 6.24: Qn4 Participants responses

For question 4, it appears that 3 participants did not answer the question therefore we have only 21 responses out of the effective 24. Based on figure 6.24, there appears to be no difference between the preferences over Manual control or Shared control but one can see that most participants tend to like Manual control or Shared control more than Automatic control.

Here are more interesting thoughts from the participants on this:

1. I prefer being in control of the movie it allows me to customize my experience. However automated control would be good if it does it right at the right times. Shared control can be frustrating as the user has to fight for control at times.”

2. “The shared control was confusing, I didn’t really know what was going on, I expected to see my hands since I allowed it but then it suddenly changed the controls and I didn’t understand it. Automated controls are easier to use, but I wasn’t able to use my legs and hands in some parts of the movies which I wanted. Manual control should be the best if I have had more time to master the gestures.”
3. “manual control make me lose interest of the movie. I just focused on playing with hand gestures.”

The first and second comments showed participants liking Manual control; the third liked it too but it became too distracting that the hand gestures alone was fun enough. The next comment is a great summary from a participant’s experience with all three types of control. It illustrates how each type has its own pros and cons which highlights the importance of what the film director wants to show his/her audience and the amount of control viewers should be given for their “involvement” in “participating” in the film.

4. “1 is Automated control because I think it is better for the movie to control it for you if the goal is to have the ”best” experience. Since if we use shared/manual control and the participant is not used to the gestures then they might miss the best parts of the movie and be too focused on the real world. 2 is shared control. Because this helps to ignite curiosity with the participant and possibly aid with their disappointments i.e. if the participant wants to see their hands, but can’t because the movie is controlling it. 3 is manual control. Because since the participant do not know much about the movie, they might miss the good parts if they suddenly decide to see the real world. But then again I think the ratings will also depend if the participant enjoys the particular movie genre. I think if the participant watches a movie genre that they DO not enjoy i.e. horror, it would be best to have manual control. But if they do not mind being scared etc. then automated control might be best.”

6.1.2 Qualitative Measures

These qualitative feedback were obtained directly from the post-experiment 1 and 2 questionnaires that every participant completed. Also more feedback were taken from

the short interview conducted with participants after their completing the whole experiment.

6.1.2.1 Session 1

This is what we discovered and learnt from experiment 1 feedback:

Many participants claimed that there is no reason for them to see their surroundings unless it plays a part in the movie such as a movie scene where the viewer and their surroundings disappear into space or something. It has to be something done on purpose by the film director for them to create a more immersive effect for their viewers. Almost all participants agree that it is good to be able to see their surroundings when they have to talk to someone briefly or take a sip of water without having to take the Oculus Rift off.

Participants wanted more realistic graphics of the user body visualization such that they could have virtual clothing that matches the movie content, so that they could fit in to the movie virtual world. Also they felt that their body right now is used for nothing. If there were interactions that they have to perform in the movie such as challenges or checkpoints to clear, they felt that it would then be more sensible to see their own hands and body. Other than that, seeing the real surroundings only hampers their movie experience. Also when there is nothing for them to interact with, some participants tend to just look around by only turning their heads and not up or down. Doing so causes them not to notice their body visualization because they do not look at themselves or raise their hands so they could see them. However if there was something they had to complete, like some form of interaction, they would need to use their hands then being able to see oneself starts to make sense.

6.1.2.2 Session 2

This is what we learnt from experiment 2 feedback: Most prefer a state change in the viewable display distance such as from nothing to just the body to the body and surroundings, when they have interact and do minor real world tasks such as drinking some water, having some food or talking to a friend briefly all without taking the Oculus Rift DK2 off. Participants find it more helpful ,faster and convenient this way. But

when interacting with the virtual movie world, a gradual change in the viewable display distance is much preferred as one would be able to adjust howsoever they like based on one's preference and where they actually are in the real world. For example, if one is near some tables and chairs then they would like to see less of them and decrease the viewable display distance, or if there is a huge open free area in front then the distance will be increased. Having this gradual change gives participants more control on what they want to see of the real world while in the movie virtual world, and minimize any distraction however they see fit.

6.2 Discussion

This section discusses the results further, and seeks to explore potential explanations for the results. In session 1, we can see that there is a significant difference in the sense of presence and level of user engagement between seeing their body and not seeing their body in the movie. However there is a balance between the live type and the animated type of movies. It is clear though that the two within subjects factors "TypeOfMovie" and "BodyShownOrNot" are not dependent on each other, and stimulates the overall experience individually.

6.2.1 Exclusion of participants

First we will explain our reasons to exclude certain participants from our effective analysis. Later it turns out that this exclusion made us explore deeper behind the effects of user embodiment and its implications in a cinematic virtual environment. This will provide guidelines of what a 360 movie virtual world should have in order to immerse viewers. Our data showed that participant 18 rated her "Animated+Body Not Present" experience as 5.4 out of 7 compared to "Animated+Body Present" 4; also her "Live+Body Not Present" experience is 6 while "Live+Body Present" is 5. Likewise, participant 22, 28 and 30 rated their "Body Present" experiences poorer than their "Body Not Present" experiences, as shown in table 6.1. However based on their post-experiment comments, if this work could be improved by including purposeful interactions, they would rate otherwise. Here are some of their comments stating so:

1. “I would love to try the experience with content that requires you to view your body/interact within the world because then I could feel more a part of the world in doing that as opposed to waving my arms without a purpose.” - participant 18’s comment demonstrate the need for creating content and interactions that encourages/requires the viewer to move, touch and experience the movie environment.
2. “if I were to see my body in the film it would be better if I were to see what I am wearing to fit in with the film. If i were wearing a body suit that could become amour like in the warcraft film then it would fit in more with the film rather than me wearing jeans and sweatshirt. Interaction-able items could be a nice touch, like if you used your phone to be a usable item in the film like a sword or a shield in the warcraft film. Maybe if your surroundings were also able to be apart of the film that could be nice, like the chair you were sitting on was designed to fit in with the decoration of what was in the film.” - participant 22’s comment highlights the possible demo experiences that can be created with this technology, and also emphasizes the participant’s feelings and sense of presence in the current system and what it could be if the system’s full potential is realised. If you see figure 6.25, here is an example of sitting on the virtual red chair, and figure 6.26 is where one could rest his/her arm on the virtual chair. Lastly figure 6.27 demonstrates the participant’s idea where one could pick up their smartphone and the system would recognise it and transform it into a weapon for the user to wield.
3. “I want to see my body in the movie if there is an animated body of myself and I can interact with the characters in the movie. That’ll be fun. I don’t like to see my surroundings in the movie because it’s so different from the movie which may distract me from watching it. But if the surrounding can be made to be part of the movie and people can use it to interact with the movie, it can be accepted.” - participant 28’s comment again includes interactions to pull viewers into the movie experience where you could see in figure 6.28 one can interact with a cute white rabbit, and how should the real surroundings of the viewers be used
4. “If the graphic resolution of my body is similar or same as the one in the film while I get to ‘dress’ up with virtual outfit, I will definitely prefer to see my body

¹StressLevelZero : Time Couch VR <http://bit.ly/1GB8eC0>

²StressLevelZero : Time Couch VR <http://bit.ly/1GB8eC0>

³Legendary VR : Warcraft Skies of Azeroth <http://bit.ly/1HDH64c>

⁴Baobab Studios Inc : Invasion! <http://bit.ly/1KwK6rB>



FIGURE 6.25: Purposeful Interactions : Sitting on the red chair¹



FIGURE 6.26: Purposeful Interactions : Resting arm on the red chair²



FIGURE 6.27: Purposeful Interactions : Using a phone as a weapon³



FIGURE 6.28: Purposeful Interactions : Interaction with virtual characters⁴

in the film. prefer not to see the surrounding: as the graphic of the surrounding seems to be different from the film and distract me from feeling being involved. I become aware of the real world whenever I see my real surrounding. if the graphics of the surroundings is improved and gets closer to that of the film, with additional functionalities where I can modify it into interaction gadgets to interact in movies, I will definitely go for it.” - participant 30’s comment too demonstrates what viewers want to see and be able to do if their body were to appear in a movie. Imagine if in figure 6.29, you could see your body dressed in Hunger Games battle armour but you can still see your hands and skin colour.

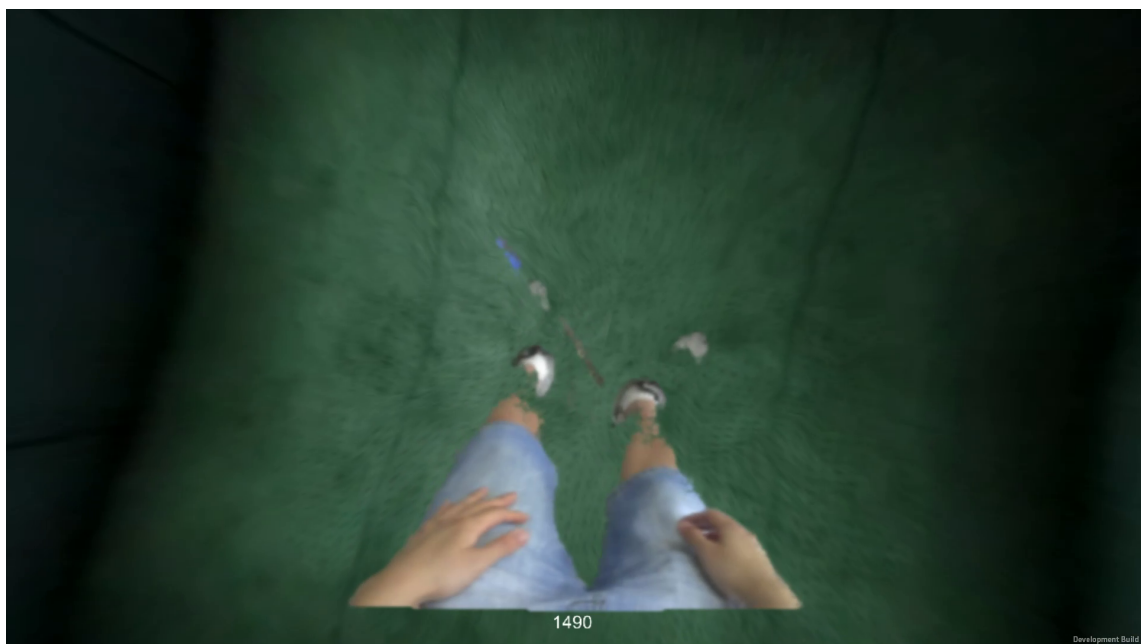


FIGURE 6.29: Purposeful Interactions : User Body in Cinematic world⁵

Thus we learnt that everyone likes to see their body in the movie but there has to be reasons for it to appear and not be displayed just because the system can. This does make sense, as if one can only see his/her own body but not use it to “physically” interact with virtual objects, it is as though there is no body actually present in the virtual movie world and only an imaginary one exists. Therefore this new insights made us question what we meant exactly by bringing their body into the movie. By removing these four participants and reducing the total to 24 participants for effective analysis, we have decided that for a convincing immersive “I am really there” experience, it is not

⁵The Hunger Games : VR Experience <http://bit.ly/1kIicwy>

enough to only give a visual representation of the real body and surroundings. It has to be more. However a side thought was what if the body and surroundings visualization can cover the entire field of view? Will that change the participants responses?

6.2.2 Further Discussion

From our quantitative and qualitative feedback for session 1, we learnt that participants generally preferred to see their body visualization in both live and animated type of movies though slightly leaning towards live movies as seen in figure 6.11. Likewise for their real surroundings visualization but now more towards live movies instead of animated as seen in figure 6.13. Regardless of the type of film, they were neutral overall towards seeing their body and surroundings in the movie but they like having the body appear more than surroundings. Please do note that responses would be different if there are interactions supported with the visualizations. Overall for session 1, participants liked the idea and concept of the system though the visualization graphics and technical aspects of the system such as field of view, display resolution of the HMD, etc could be improved.

The field of view of the visualization was a problem but we could not fix it due to a hardware limitation with the RGB-D DS325 cameras. One participant perfectly put it this way: “Having to look down to see your body wasn’t particularly natural and didn’t really improve the experience that much. It would be better to simply be peripherally aware of your body, as this is what you are naturally used to. A larger field of view could help with this.” Also the playback was not at a high frame-rate and some could notice some judder at times, this is another technical aspect of the system to deal with. Many mentioned that now with a body they felt like moving around but they could not. Some wanted to “push the car that was parked at the side” and “punch the soldier” with their hands and feet. Some wanted to talk to the character since they are now “physically in” the movie. The list of improvements could go on and more if we were to include haptic feedback which a few participants suggested; but at this stage including interactions with the movie through the user body and surroundings visualization, and improving the system hardware are most important.

Before participants completed session 2, they indicated that they were more keen on controlling their body visualization themselves and were neutral to have the movie system to do it for them, though they would actually like it if both parties could share this control. Based on session 2 results and feedback, we see that people were mostly neutral slightly leaning to agree towards using hand gestures to control the appearing/blending/viewable distance of the body and surrounding visualizations. Refer to figure 6.23. One of their reasons to not liking hand gestures was that they could not use their hands to touch or hit objects in the movie (though since there is no interaction supported, this is more of an imaginary touch and hit) as the system might pick some of their hand poses as gestures and make changes to their visualization.

Participants were divided between Shared and Manual control with an equal number of people selecting either. Automated control was the least desired. The feedback, before and after session 2, from the participants is exactly the same. Many had fun with the gestures but thought it could be more intuitive and better improved in sensitivity. Again the field of view issue, one participant put it this way: “Seeing half of your arm was kind of weird and didn’t particular help the experience.”, another: “Dislike the square box of the real body showing in the midst of the virtual world, it reminds me that i am in virtual world when I want to go back to reality.” One participant’s comment highlighted the need for interactions in the movie to prompt viewers to use their gestures or do things: “I feel like I didn’t need to use my hands and stuff in a movie situation. A more interactive movie would be better.”

Remember the movies mentioned earlier that has the computer generated bodies of the first-person main character in them such that if the viewer looks down they would see an filmed or animated body there. If we were able to take this clothing and put it onto the viewer’s actual real body visualization, it would bring immersion to a whole new level. This is what our participants mentioned they wanted during the think-aloud interviews after the experiment. All felt that if this virtual clothing feature was to be introduced, they would definitely think they would be more engaged and into the movie than without it. This is different from our current system where viewers can see their bodies but they cannot use them to do anything. I think this is the huge reason why some participants preferred not to see their bodies in the movie. Seeing their bodies in the movie and not being able to do anything with it is more of a distraction, but once you can interact with your body present then brings the immersion to a whole new scale.

This is hugely why we think that this analysis should be done again once our current system has the features capable of supporting body interactions and customizations. It is because of these responses from all participants including the excluded 4 participants, that we then could exclude the 4 participants who rated hugely negative on the body visualization. Doing so, allowed us to receive significant results for both sense of presence and engagement.

6.3 Participants-feedback-inspired Future Improvements

The real surroundings visualization obstructing the movie scene was a noticeable issue, and participants had to take some time to gesture for the surroundings to disappear in order to focus on the movie again. As a result with inspiration from a participant's comment, we proposed that during the movie, state change should be used to transition from nothing to body visualization to body and surroundings visualization to back to nothing. This would be faster and easier, but when needing to tend to real world issues such as take a drink of water or talk to a friend briefly or answer the phone all without taking the Oculus Rift off, the gradual transitions would be better as one can adjust the blending/viewable distance through gesture controls as needed.

The lack of sensitivity of the gesture recognition in session 2 made some participants tired with their arms. One suggested since the movie is an individual experience, it would be nice to have a rewind, pause and forward control of the movie.

Here are some quoted participants thoughts on what to improve on that are worth to think about:

1. "Choosing clothing appropriate to the virtual environment will be a welcome addition."

As mentioned earlier above about the clothing of filmed or animated main characters in first-person 360 movies, if these clothing could be overlayed on top of the viewer's body visualization, it would greatly increase the immersion experienced.

2. "The use of the gestures results in the appearance of the real world in a 'window' which obscures the virtual scene making it hard to follow and is distracting. The surroundings between 1000 - 1400mm where only certain things such hands are

visible is fine since at this distance it appears to be part of the virtual environment. A wider field of view for the camera is acceptable as long as the film pauses automatically when this is engaged.”

3. “I think making use of the oculus rift without have to take it off is a good idea with the gestures thing that I can use to increase the visible real world surroundings in order to do my trivial real world tasks because it can be a major distraction taking the oculus off while watching the movie.”

The trivial tasks this participant mentioned above are such as drinking water, finding and answering a phone, etc. Basically anything that requires little attention, focus and brief moments using the eyes. Things like reading a text on the phone is definitely not recommended to do as the eye strain is immense as the clarity of the RGB-D camera may not be good enough to be readable.

4. “I would prefer if the automatic control was done intentionally by the movie director because the director would want the audience to see from his/her perspective when watching the movie. It would also be good if for example, I wanted to grab something without removing the oculus, I want to be able to control it so that I can see enough of my real surroundings to find it.”

This participant above is suggesting that the movie be played with automatic control, but if he/she wants to do some trivial real world tasks such as grabbing a drink, they want to have manual control to self-adjust what they can see without taking the Oculus Rift off.

5. “If I can’t see the real surrounding while watching the movie, it would be hard for me to find the most comfortable position or place to watch it as I would have to take the device off first and that would interrupt my movie experience. However if I can control what i can see in the movie from the real surroundings, I can always control it so that I could find let’s say a chair to sit in when my legs are tired after watching it while standing up.”

We believe the system was designed in the right direction. For a portable and high-performance small RGB-D camera, DS325 was one of state-of-the-art devices at the start of 2015. With more time the system software can be further improved, better utilising computing resources to increase the frame-rate and quality of the rendering.

Features such as gesture-based interactions and activity with virtual objects can be built on top of the SoftKinetic iisuTM platform to improve the natural harmony between the real and virtual worlds. Gesture recognition can be further configured and be made more sensitive. More testings have to be done to get this right. Our system was mostly limited by the field of view of the DS325 camera. With the visualization working fine using the DS325, one can attempt using two or more DS325 cameras and explore its performance.

6.4 Conclusions

Our system design grew a lot from the focus group, the TedX conference demo to the actual experiment. Our main research goals started from just looking into whether there is difference in the viewer's sense of presence when their body appears and disappears into the movie scene; also if there is a difference with various types of interfaces used to control this appearing/blending. Then we added if not only the sense of presence would be affected but also the viewer's engagement. With much research and choosing the real type of movies to play, it was interesting to see if the realism of the movie would affect the viewer's sense of presence and engagement too.

With 28 participants, we were unable to find any significant results for all three hypotheses. However when analyzing deeper into the sub-scales of IPQ and Engagement tools, we found a few participants with huge negative ratings with the "BodyShownOrNot" factor. This was not a surprise as this was discovered during the post-experiment interview sessions with participants. Almost all participants agreed that seeing their body visualization and real surroundings will give a greater sense of presence and engagement when there are more meaningful features implemented, such as putting on virtual clothing on the body, and using their real surroundings as tools to interact with movie objects and characters. With qualitative feedback to back up these claims, we excluded those few participants to reduce the total to 24 participants. As a result, we found significant differences in the "BodyShownOrNot" main effect for both IPQ and Engagement; but the "TypeOfMovie" main effect and interaction effect were not significant.

It is true that based on what our system can only provide right now, this analysis results is short-lived and should be analysed again once all features are included into the

system. Following expectations and suggestions from participants feedback, they like the whole idea and concept of the user body visualization and real surroundings blending. However for this system to deliver and satisfy their hunger for immersive entertainment, it needs to have a decent working example which we could only partly produce. A fully deliverable system would have a movie specially catered for immersive 360 panoramic viewing with high-quality surround point-sources audio; also it should have high quality graphics rendering of the viewer's body capable of removing and putting on virtual clothing elements; furthermore it shall support and recognise certain key surroundings around the viewer, which can be used as tools of interaction in the movie to interact with the characters or even virtual objects.

Chapter 7

Conclusion and Future Work

7.1 Conclusion

Our ultimate aim of this research investigation is to recreate experiences by blending reality and virtual reality. Interacting and truly engaging with the movie content and characters requires you first “being there” and that usually means being able seeing yourself “there”. VR technology is designed to serve the needs of the users’ sensory channels: our eyes, ears, hands, and so forth. Each channel “expects” information in a certain way. The needs and demands of our senses play a major role in determining the value, quality, and utility of a VR system component.

As discussed in the introduction, our research aimed to contribute to the field of immersive movie entertainment with HMDs by exploring ways of augmenting the virtual space with real-time 3D visualization of the hands and body of the user. This allows them to perceive themselves being physically present in a virtual environment. We are also interested in how this visualization can be utilised to create a transporting effect on users from reality to virtuality, how users would react if this transitioning is controlled manually, automatically or shared between the system, whether this control of fading in and out of mixed reality should be given an interface, what kind of interface would be best in a movie scenario and how would it work?

Back then 360 movies/videos were only starting to be popular, and there were no clear guidelines or resources on how to properly design these experiences or even film them. One common doubt we found in all 360 movies was should they be jam-packed with

action, leaving the viewer free to explore on their own, or should they be baby-sitting the viewer guiding them through every scene and specifically showing them what to look at. On our own, we studied what short movies and trailers were out there, and filtered them ourselves with a set requirements list we made and were interested in. After a focus group and brainstorming session, a prototype system was created just in time for the TEDx conference in Christchurch, New Zealand. Valuable feedback and suggestions about the movie, factors regarding the movie, and the system's performance most specifically the head-shaking feature were obtained from the participants. The head shaking interface was ultimately removed due to its insensitivity and tendency to increase nauseousness from the vigorous shaking. Thus the hand gestures interface was designed as its successor for viewers to move seamlessly between the real and virtual world. Out of all the characteristics of the 360 movies shown in TEDx, participants feedback pointed out that the difference of either being a live-action movie or animated movie stood out the most. A live-action movie would be defined roughly as movies with real human actors/actresses based on a real world scenario though they can still have visual effects (VFX) and computer graphics (CG) included. One example would be *The Lord Of The Rings*. An animated movie would be completely CG generated creating a cartoony or animation feel such as *Despicable Me*.

Our user experiment consisted of two sessions, the first explored the effects of two factors with two levels each, that is, two levels of "Body present or not present", and two levels of "Live movie or animated movie". Results showed a significant difference in sense of presence and level of user engagement in the user body visualization factor though there were few who felt strongly that visualization was not enough and should come with meaningful purposeful interactions. In summary, we derived qualitative feedback such as : "I want to do something with my body when I am able to see it."

There has been a lot of research into user embodiment in VE, however not when a 360 movie is played, enveloping the viewer into a whole new virtual world. The fact that this is more than a virtual environment but now with a story, action and emotional scenes, influences differently how viewers feel present and engaged. One could suggest that it was because they felt present there that they wanted to interact and engage with the movie, but once they realised that they could not due to the technical limitations of the system, they are put off and the immersion diminishes. To answer our first and third research questions in section 3.1.3, blending the user body and accompanying

environment into the movie scene does improve the user experience. However we learnt that for 360 cinematic VE, to be able to see is insufficient, as activity and interactions must be encouraged for the user experience and story from the movie to harmonise. Separating either of them makes it feel something is missing, causing the high levels of immersion at the start of the experience to fade away, some more than others.

Results on “Live movie or animated movie” was insignificant. However when asked verbally, including the participants from TEDx, almost all mentioned that in live type of movies it is more appropriate to see their user body visualization. It could be that what they thought does not reflect what they actually felt. We think that the limited FOV of the real body and surroundings visualization could have played a part in negating any differences in presence and engagement.

In the second session, we were not only interested in user body visualization but also surroundings. Research was conducted between three types of control over the user body and surroundings visualization blending in and out of the real and virtual worlds: Manual control (by viewer only), Automated control (by system only), Shared control (by both viewer and system). Hand gesture based interactions were also used to control the visualization blending, and the feedback from participants on what they thought of it, such as its appropriateness in a movie scenario and the techniques used, were important. Results reflected that participants were mostly neutral to slightly positive towards using hand gestures. There were many suggestions on what the right interface should be, but all had one thing in common: it depends largely on the movie and its content. If the movie aims to be more like a game then more controllability and accuracy is needed such as buttons. Also if it is action packed then certain hand gestures cannot be used. But if a more dwelling and introspective experience is pursued, simple controls like the current hand gestures is sufficient. Here we have answered our second research question in section 3.1.3, regarding the best way for viewers to control their transitions between the real and virtual worlds. Results on type of control came back with participants undecided between manual and shared control, with automated being the least preferred. Again this depends on what type of 360 movie is being watched. The director of the movie knows his/her film best therefore would know which scenes are best to show the viewer’s body and when not to. Also the blending from real to virtual can be intentionally utilised in the movie, such as being beamed from the real physical world to virtual outer space.

A huge amount of emphasis on what is the right interfaces, techniques and design used in the system is placed on the kind of 360 movie content. This is extremely important for movie directors of the future. A more detailed and rigorous research has to be done to fully understand the entirety of 360 movies and its implications.

7.1.1 System with a focus on Home Entertainment

When this work started, there was a lot of research done on using HMDs to view visualizations of data/objects which sometimes can be touched and changed through hand or tool interactions; for example using HMDs as a platform to create immersive environments such as in museums to make their history stories more appealing [52]. However they were all limited by a fixed huge set-up with trackers and other technical equipment needed which an everyday household would not have. With a side goal of making the whole system portable and easy to set-up where you would just have to plug in a few wires and power plugs in your home desktop PC, relying on just the camera to do the tracking would be a whole game changer. Being portable comes with huge costs which we saw with the many technical and software issues. However the current 2016 recommended desktop specifications/hardware standards to support the intensive high quality rendering HMDs such as the Oculus Rift, HTC Vive, and Sony Morpheus, are quite intensive with at least a NVIDIA GTX 970 / AMD R9 290 equivalent or greater video card, Intel i5-4590 equivalent or greater CPU core, and 8GB+ RAM. With these hardware supported, I believe with our current set-up using the SoftKinetic DS325 camera alone can be improved. Software-wise, the movie playback graphics quality can definitely be improved through heterogeneous computing such as compute shaders and hardware acceleration or other solutions. Visualization of the body and surroundings can be re-designed to incorporate more than just particle systems. Calibration and recognition walk-throughs before the movie can be considered to set up the viewer's body and his/her surroundings in order for the system to use them as interactions during the movie. Just as similar to newer desktop specifications, if there is a better lightweight portable RGB-D camera than the DS325, that will hugely improve the entire system from the FOV to performance.

7.1.2 Future of 360 Film-making

In the past decade, we have seen the shift in the industry of film and game creation steadily coming closer. Game artists and film animators crossing over to collaborate is no longer seen as unthinkable. Just recently, Epic Games showcased a stunningly made digital short film “A Boy and His Kite”. The production was the perfect unity of the two worlds, and gave the impression of a high-end feature film. This two minute film debuted at the Game Developers Conference some time in 2015. The story also tells the battle of merging game and film creation as technologists and artists craved for solutions and tools to do so. The Unreal 4 engine is this tool. In relation to our work, we too hope that this augmented virtuality system we have created will be the start of the merge between film, games and our real physical world. With 360 cinematic content different from traditional 2D films, the human element is important which is why we believe 360 movie directors would need this system technology running in real time as they shoot their films. High quality post-production generated on high-end machines employing state-of-the-art techniques will always be needed to create great amazing aesthetic imagery. But the production/shooting of the film will be larger in scale in terms of lighting, set-ups and acting, which is why it is even more crucial for directors to be able to get into the perspective of the viewer through our system and preview the shots on site. This re-affirms our goal towards the portability of the cinematic VE system.

7.2 Contribution

1. Exploring 360 movies and its implications to user embodiment

There is a lot that we do not understand about the uprising popular 360 movies nor the effects of placing ourselves in 360 movie virtual environments. Graphics quality-wise, the constantly changing moving viewpoints in 360 movies raises a lot of challenges to image-based lighting and etc for realistic visualizations. Putting aside these difficulties, we want to understand people’s reactions to seeing themselves in a constantly changing environment.

We broke down 360 movies into their separate characteristics (genres, view perspective, etc) and observed its effects on presence and engagement. Based on

participants feedback, whether the movie is live or animated mattered the most to them. Though we had no significant differences in sense of presence and user engagement, we highly suspect it might be different if the FOV of the visualization is expanded. Also since we were using movies obtained from YouTube and not filmed on our own, we could not control every characteristic of the movie which could influence our investigation.

2. Incorporating a view of the user's body through visualization into an immersive 360 movie

A lot of work has been done on user embodiment in virtual environments but not with the 360 movie as the virtual world. The fact that it has a story and a deeper emotional connection with users influences differently on how viewers feel present and engaged in the virtual environment. Our research showed a significant difference in the sense of presence and level of user engagement, though some participants felt that seeing their own body was not enough and should come with meaningful purposeful interactions. This has to be explored deeper but is understandable and insightful in how a VE with a story behind it can be so different from a traditional VE.

3. Blending views of both the user's body and surroundings, through visualization controlled by different designs, into an immersive 360 movie

Based on an idea of a dream, fading from the real physical world to an imaginary virtual world, we explored people's preferences on how much control they would like to have over the transition. There are three types of control: Manual control (by the viewer only), Automated control (by the system only), Shared control (between the viewer and system). Results showed viewers are split between shared and manual control, with automated being the least preferred. Participant feedbacks commonly address the importance of the movie director's influence, such as if one of the best scenes requires the viewer to see his/her body, then the system should have total control since the director has intentionally done this. Furthermore we believe this type of control can be generalised for any other interactions between a viewer and the system in a cinematic VE.

4. Using hand gestures to control the visualization blending of the user's body and surroundings, in an immersive 360 movie

Participants were mostly neutral to slightly positive in using hand gestures as the interface to interact with the controls over the visualization blending. Participant feedbacks commonly address the type of 360 movie content is played. If the movie has more game elements then more controllability and accuracy is needed such as buttons. But if the movie game is action-packed and requires the usage of hands then certain hand gestures cannot be used. If the movie is a more dwelling and introspective experience, simple controls like the currently implemented hand gestures are sufficient.

5. Explore how visualization and its control can meet the practicalities of a home entertainment system

We found that one could use the controls over the visualization blending as a way to stay or switch connection to the real physical world or virtual world. A few reason why people would want to are 1) to have a drink or eat snacks while watching the movie 2) briefly talk with friends sitting beside you 3) find your smart-phone to answer a phone call 4) most importantly we can do this without taking the HMD off which is known to significantly break immersion.

Participants loved the idea and based on their feedback, the system's usability can be further improved. The currently implemented blending in and out of the virtual and real worlds is great when needing to do real physical tasks as mentioned above. But within the cinematic VE, a modes approach should be used such as one mode would have only the virtual world viewable, next mode would be with user body visualization, another would have user body and surroundings visualization, last mode would be just the real physical world viewable. The swiping left and right gestures would be the best for swiping between modes so that viewers can easily switch to their desired visualization in one quick instant.

6. Proposal of a home entertainment system with user body visualization capabilities

This system design is one portable lightweight solution to the complex registration and calibration set-ups needed for immersive cinematic experiences. Depth keying is needed for visualization blending between virtual and real world, thus at least one depth camera is needed. RGB-D Camera/s to support a large FOV of the visualization, ideally as large as the HMD used such as Oculus Rift. The usability

design for fading/blending while engaged in the cinematic VE, is a swipe-between-modes approach utilising both “SWIPE LEFT” and “SWIPE RIGHT” SoftKinetic iisuTM supported gestures. Usability design for fading/blending while not engaged in the cinematic VR to do minor real-world tasks without taking HMD off, is a gradual-increasing-decreasing approach exactly how our current system works, utilising the “BIG 5”, “THUMB UP”, AND “THUMB DOWN” SoftKinetic iisuTM supported gestures.

7.3 Future Work

This work is a foundation upon which we can build a dialogue about user body altering and visualization, encouraged interactions, contact with virtual characters, and further improvements to the system both hardware and software-wise.

7.3.1 Application with Gestures

Supported with Gibson’s [53] ecological approach to visual perception, Flach and Holden [54] states that the technology’s affordances for action determines the sense of presence in VEs, e.g., the bodily practices of which the avatar is capable. Affordances are resources reveals the user’s desire to take certain actions[55]. Thus they are dependent on an individual’s intention and the cultural context in which it occurs. It is known that users are more present in a perceptually poor environment like text-based virtual worlds, in which they can act and interact in a way that their goals are realized, than in a 3D graphical virtual world in which they cannot [56] [18].

Applying the notion of embodied cognition [57], which argues that environments are conceptualized in terms of bodily experience, Schubert [58] developed an embodied presence model[18]. They determined presence as a subjective experience that depends on a multi-dimensional process through which users understand and construct virtual worlds. This model is based on these principles:

- **Spatial presence**

which represents the mental image of one’s own body being in and interacting with the virtual

- **Involvement**

which is the attentional component that signifies the user's interest and engagement with the content of the virtual environment

- **Judgement of the virtual environment's realness**

which relates to the user's actions and interactions within the virtual world

Imagine playing a movie version of Clash of clans in VR, where this mobile tablet game is one of the most popular games in the app market. From early on, they had a clear goal for their gestures control targeting the best way of getting the scrolling and tapping to work. A good gestures system implemented easily gets viewers involved with their body present, engaging with the movie and its action affordances. In Clash of clans VR, it is possible to use the simple gestures of tapping as shown in figure 7.1 or pushing forward to kill enemies in your line of sight with a first-person perspective. Or one could use their smart-phone as shown in figure 7.2 and the system camera will recognise it and visually transform it into a sword for you to wield in the VE.



FIGURE 7.1: Clash of clans VR - Tapping enemy to die¹

A great movie example to demonstrate all of the potential system features with well supported gestures would be a war action movie. The viewer will be welcomed with

¹Clash of Clans 360 VR <https://youtu.be/wczdECcRw0>

²Clash of Clans 360 VR <https://youtu.be/wczdECcRw0>



FIGURE 7.2: Clash of clans VR - Smart-phone transform into virtual sword²

a title screen before proceeding to the locker room. If the viewer were to look down, he/she could see their own body visualization. When they walk to their locker, they will be provided with a few body armour set selections that they could wear; gestures such as swiping left or right and pushing forward to select can be use to pick their choice. Once selected, they could look down to see their own body now fitted with the body armour set of their choice but distinctive features like some of the skin they can see through the armour and their fingers are of their own. Next they will be proceeded to a war zone where they are dropped off by a bungee. If the viewer takes out his/her smartphone, the DS325 would recognise it from its computer vision database and turn it visually into a gun which the viewer can use to shoot at enemies.

7.3.2 Cinematic VE Experience - individual or group?

Another aspect of the system that we can work on is based on the idea of whether this cinematic VE experience is going to be individual or group. In our system implemented, we assumed that it will be an individual personal experience. Since the viewer is watching the movie independently, we could provide additional controls to rewind, pause, and forward the 360 movie. Pausing the movie could mean that the viewer might want to take a break or possibly grab something to eat, this is when the system could switch from the swipe-between-modes controls to the gradual-increasing-decreasing controls.

However one could argue that viewers might pause the movie not to do minor real world tasks but to look around the 360 movie especially when a lot of action is taking place in the scene.

The rewind-pause-forward controls is useful for an individual personal cinematic experience; but how should its interactions be designed? Would we use hand gestures? And what implications of pausing the movie do we follow? Should the swipe-between-modes controls be changed to the gradual-increasing-decreasing controls when the movie is paused? Or should we literally pause the movie and give the viewer the option to switch the controls if they want to perform minor real world tasks? There is much to research and understand about supporting a rich individual 360 movie virtual world experience. Not to mention the unexplored multi-dimensional world of multi-viewers cinematic VE, having more than one viewer experience a 360 movie together at the same time, brings its new form of challenges and concepts that we do not understand.

7.3.3 Extending FOV

If you look at figure 4.16, you could see that the real world being shown is limited by the field of view of the DS325 camera. Imagine expanding this region by placing two or more of this “windows” side by side, that is, using two or more DS325 cameras covering almost the whole noticeable area on the HMD display screen. This would help us to answer the question mentioned earlier on when we had to remove the 4 participants from our effective analysis. It would be interesting to see that if with this current visualization without any purposeful interactions supported but now en-covering the entire field of view of the HMD, would make a difference in the sense of presence and level of user engagement.

7.3.4 Head Mounted Displays

HMDs are an almost ancient tech which have been given a reboot in the past few years as computers get more powerful, and video games steering the way to greater visual graphical perfection. There are actually three different classifications of HMDs. First would be like the Google cardboard, a head mount with a standard LCD screen such as a smartphone to display images and videos.

Next would be augmented reality, a live or indirect view of the physical real world overlaid or augmented with projected images on top of a pair of see-through glasses or googles. Google's Glass³ gave the general public their first taste of AR but the project was shut down some time ago. Right now the strong contenders are Microsoft's HoloLens⁴ and Magic Leap's⁵ new secret device that they have not named. Right now there is an enormous amount of hype around them and what they can do concept-wise, however right now they are not ready. As a matter of fact, at this point in time of writing it is known that the first version of the HoloLens device has a similar problem as us: limited field of view. You are only able to see virtual content in a small area of about 35-40 degrees⁶ horizontally which is actually smaller than the DS325 camera. It will be very interesting to see what Microsoft's and Magic Leap's final prototypes would be like in the future.

Lastly we have virtual reality, replicating a real physical environment by stimulating the sight and auditory senses, and allowing the user to interact in that virtual world. Devices range from the well-known Oculus Rift⁷, HTC Vive⁸, and Sony PlayStation VR⁹. At this time, other than Oculus Rift releasing its DK1, DK2, and soon to come Rift, Vive has also been just recently brought into the hands of users. What is even more exciting is these new hardware also comes with their own controllers built to be used alongside with their HMDs counterparts. For the Rift, you will have a pair of Oculus Touch, lightweight, wireless and handheld motion controllers. They are fully tracked in 3D space by the Constellation system. With the HTC Vive, dual SteamVR single-handed controllers are used as inputs. One thing in common between them is that though they are represented in VR as real physical hands, in reality they are still obtrusive as controllers. The Vive package also includes two Lighthouse base stations that sweep structured light lasers within a space. The front-facing cameras of the Vive then is able to identify any static or moving objects which allows the software to warn users of any close by obstacles.

³Google : Glass <https://www.google.com/glass/start/>

⁴Microsoft : Hololens <http://www.microsoft.com/microsoft-hololens/en-us>

⁵Magic Leap www.magicleap.com

⁶Wired : The HoloLens isn't as great as you think <http://bit.ly/1FH62vA>

⁷Oculus : Rift www.oculus.com

⁸HTC and Valve : Vive <http://www.htcvive.com/>

⁹Sony : PlayStation VR <https://www.playstation.com/en-nz/explore/ps4/features/playstation-vr/>

7.3.5 Unreal Engine 4

Unreal Engine 4 similarly to Unity, is a complete suite of game development tools for building every aspect of your project. Every developer has access to the complete C++ engine and editor source code. Having the full source code enables the power to customize your application, making it easier for debugging. Its blueprint visual scripting is renowned for rapidly prototyping and building without a single touch of programming. Additionally like Unity, it is multiplatform supported so that developers just need one engine, one workflow and be able to meet the needs of many.

Because Unreal Engine is designed for demanding applications such as AAA games and photoreal visualization, it meets the requirements for immersive VR experiences and provides a solid foundation to build upon. Similarly to Unity, it has frequent updates to its support for all the latest VR hardware such as Oculus Rift.

The Unreal Engine was initially planned to be brought it during later stages so that we could port out work from Unity to Unreal which is well-known for its stunning and amazing realistic graphics. But time caught up and there was not much left we could afford to learn Unreal Engine 4 and re-build the experiment application. Therefore we stuck with Unity. A remake using Unreal Engine 4 could be a future work.

7.3.6 Light Fields - The Future of VR-AR-MR -Light Fields

The world of virtual reality (VR) encompasses immersive head gear experiences such as Oculus Rift and also extends to augmented reality (AR) and 'Mixed Reality'(MR), which is most identified with Magic Leap and Microsoft's HoloLens. The difference is best illustrated by the now discontinued Google Glass and Magic Leap's as yet unreleased newer devices. The 'screen' of data on Google Glass moves with your head. It is fixed in relation to your eye. The data or overlay of information in the Magic Leap headset however will track with the world. This second approach allows for example a digital chess board to 'sit' on a table in front of you and stay fixed relative to the desk as your head moves very much unlike the Google Glass display.

When you are watching a stereo movie with normal cinema passive glasses your eyes will converge at a point in space indicating depth, but the perception of depth of an actor

is depending with where you are focused. In other words, no matter how far in front or behind the screen you are, your brain converges the stereo pair into thinking the same action is taking place. The way for this to work is if the gear used not only does the normal stereo trick of correct stereo convergence but also changes the focus to match that of your view of the action.



FIGURE 7.3: Light Fields Concept Idea - Magic Leap and Weta Workshop¹⁰

With Light Fields technology, this is another path which VR/AR/MR cinema can take. This would no longer need to use a 360 spherical panorama film as the movie content will be set placed relative to the viewer. This would solve the problem of having multiple perspectives and viewpoints in the movie and having to cater to each of them. For this the movie virtual world can be created and the viewer be placed in a default position in that world. They would then be free to explore and experience the multiple events that would occur throughout the movie.

7.3.7 NextVR

NextVR¹¹ built their own light field solution based on live broadcast technology, extracting depth information, and keep it out so it is not subject to depth decimation before rendered as part of the final presentation. This meant that all the rigs at NextVR have

¹⁰Magic Leap www.magicleap.com

¹¹NextVR www.nextvr.com

had feature detection and depth information generated as part of their pipeline. The depth map is used to create a simple mesh of the environment and they use that mesh basically to build a representation of the world. Then texture mapping is applied to that virtual world to give it its look. NextVR is focused on a few key VR areas but a main one is live VR coverage of say a sporting event. It is in this context that NextVR uses technology including light fields to produce a more interactive experience. One example is a depth map of a basketball court from the position of the intended live coverage, figure 7.4. Onto this 3D stage they effectively project the real time captured event - thus providing more parallax. It is in making this stage that a light field could be used, but the team also use LIDAR and other non-realtime tools.

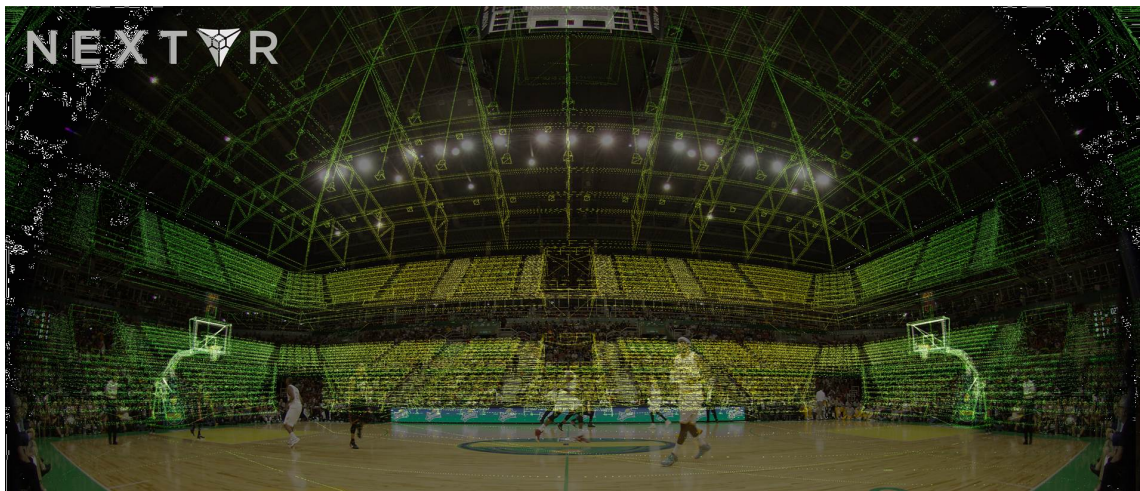


FIGURE 7.4: Light Fields Concept Idea - NextVR¹²

7.3.8 Future User Studies

With the combination of upcoming and progressing new technologies, we would be interested to conduct user studies in bringing different content domains across and together. More work will definitely be done with regards to cinematic augmented virtuality and augmented reality. But into different content types, one common well known example would be, how would bringing your own body into a video game be like? As mentioned earlier on, a game would require more accuracy and definitive controls. Interactions are much larger-scale and abundant as compared to a movie which is less defined and more versatile in concept. A more abstract content style would be for example combining musical with visuals, exploring the concepts of realism with music and sound. We do still

¹²NextVR www.nextvr.com

want to represent the viewer with their own body but how should this be represented to fit within the indefinite context?

Appendix A

Appendix A: Information Sheet and Consent Form

The following Human Ethics Approval letter was given by the Human Ethics Committee, University of Canterbury.

The following Participant Information Sheet and Consent Form were given to participants before the study.

HUMAN ETHICS COMMITTEE

Secretary, Lynda Griffioen
Email: human-ethics@canterbury.ac.nz

Ref: HEC 2015/87/LR

25 November 2015

Joshua Chen
HIT Lab NZ
UNIVERSITY OF CANTERBURY

Dear Joshua

Thank you for forwarding your low risk application to the Human Ethics Committee for the research proposal titled “Augmenting immersive cinematic experience/scene with user body visualisation”.

I am pleased to advise that the application has been reviewed and approved.

With best wishes for your project.

Yours sincerely



Lindsey MacDonald
Chair, Human Ethics Committee

PARTICIPANT INFORMATION SHEET

RESEARCH STUDY: Augmenting Immersive Cinematic Experience/Scene with User Body Visualisation

RESEARCHERS: Joshua Chen, Dr. Gun Lee, Dr. Christoph Bartneck, Prof. Mark Billinghurst

INTRODUCTION

You are invited to take part in a cinematic experience research study. Before you decide to be part of this study, you need to understand the risks and benefits. This information sheet provides information about the research study. The researcher will be available to answer your questions and provide further explanations. If you agree to take part in the research study, you will be asked to sign the consent form.

PURPOSE

The purpose of this study is to identify the significance of blending/combining what you can see in the real world while playing cinematic content through a head-mounted display (HMD) such as the Oculus Rift.

PROCEDURE

The study will follow the procedure outlined as below:

- 1. The participant reads information sheet and signs the consent form.**
- 2. The participant answers to a questionnaire individually about demographic information and his/her previous experience with using HMD.**
- 3. The researcher explains the study setup and experimental tasks for the participant to perform during the study.**
- 4. The participant performs the experimental tasks including:**
 - Wearing the Oculus Rift HMD and pilot testing with mounted camera**
 - Perform the given task, which may include watching a short movie, viewing one's own body blended into the movie, viewing the real world surroundings blended into the movie, and controlling when/how this blending is performed.**
 - Rate personal thoughts and feelings based on each task by answering a questionnaire.**
- 5. The participant answers to a questionnaire asking for feedback on the overall study.**
- 6. The participant will be asked a few questions for a debriefing interview.**

The whole procedure will take approximately 50 minutes.

RISKS/DISCOMFORTS

Risks are dependent on individuals in this study. As you will be wearing a head-mounted display (HMD) a number of times, you might feel some form of nausea or giddiness. Please voice out when you feel even the slightest of nausea so that the researcher could assist you in safely removing the head-mounted display. Being still with eyes closed could help further prevent nausea coming in rapidly.

CONFIDENTIALITY

All data obtained from participants will be kept confidential. In publications (e.g. Thesis, a public document which will be available through the UC Library), we will mainly report the results in an aggregate format: reporting only combined results and never reporting individual ones. In case of reporting quotes of the participants from the questionnaires, we will keep the source anonymous. Video of the experiment will be recorded for analysis purposes. The recorded video will be only of what participants can see through the Oculus Rift head mounted display. There will be no recordings of the participant's face, maintaining the anonymity of the participants. All recordings will be concealed, and none other than the researchers will have access to them. The data will be kept securely for a minimum period of 5 years and will be destroyed after completion of the research project.

PARTICIPATION

Participation in this research study is completely voluntary. You have the right to withdraw at any time or refuse to participate entirely.

COMPENSATION

Upon completion of participation in the study, the participant will receive a \$5 gift voucher.

APPROVAL OF THIS STUDY

This study has been reviewed and approved by the Human Interface Technology (HIT Lab NZ) and the University of Canterbury Human Ethics Committee Low Risk Approval process.

QUESTIONS

If you have questions regarding this study, please contact the researchers at the HIT Lab NZ:

Joshua Chen (joshua.chen@pg.canterbury.ac.nz)

Dr. Gun Lee (gun.lee@canterbury.ac.nz)

Dr. Christoph Bartneck (christoph.bartneck@canterbury.ac.nz)

Prof. Mark Billinghurst (mark.billinghurst@canterbury.ac.nz)

Please take this information sheet with you when you leave.

PARTICIPANT CONSENT FORM

RESEARCH STUDY: Augmenting Immersive Cinematic Experience/Scene with User Body Visualisation

RESEARCHER: Joshua Chen (joshua.chen@pg.canterbury.ac.nz)

SUPERVISORS: Dr. Gun Lee (gun.lee@canterbury.ac.nz), Dr. Christoph Bartneck (christoph.bartneck@canterbury.ac.nz), Prof. Mark Billinghurst (mark.billinghurst@canterbury.ac.nz)

I have been given a full explanation of this project and have had the opportunity to ask questions. I understand what is required of me if I agree to take part in the research.

I understand that participation is voluntary and I may withdraw at any time without penalty. Withdrawal of participation will also include the withdrawal of any information I have provided should this remain practically achievable.

I understand that any information or opinions I provide will be kept confidential to the researcher and the administrators of the research project and that any published or reported results will not identify the participants. I understand that a thesis is a public document and will be available through the UC Library.

I understand that whatever I can see through the head-mounted display will be recorded in video form.

I understand that all data collected for the study will be kept in locked and secure facilities and/or in password protected electronic form and will be destroyed after five years.

I understand the risks associated with taking part and how they will be managed.

I understand that I am able to receive a report on the findings of the study by contacting the researcher at the conclusion of the project.

I understand that I can contact the researchers or supervisors listed above for further information. If I have any complaints, I can contact the Chair of the University of Canterbury Human Ethics Committee, Private Bag 4800, Christchurch (human-ethics@canterbury.ac.nz)



By signing below, I agree to participate in this research project, and I authorize recordings or other materials taken from this study used for scientific purposes, and I consent to publication of the results of the study.

Participant (Print name)

Signature

Date

Appendix B

Appendix B: Questionnaires

The following TEDx Questionnaire was given to participants who tried the TEDx demo.

The other following questionnaires were given to participants during the user experiment. Each participant filled out one pre-experiment questionnaire, four per-condition questionnaires for session 1, one post-session-2 questionnaire, and one post-session-2 questionnaire.

Content Watched: _____

Age: _____ years old

Please **check on one box** next to the option you would like to choose.

Gender: ☐ Male ☐ Female ☐ Don't want to specify

Have you **used Head-Mounted Displays before?**

(For example, Oculus Rift, Samsung Gear VR, or HTC Vive)

- ☐ Never
- ☐ Only few times (_____ times)
- ☐ Yearly (_____ times a year)
- ☐ Monthly (_____ times a month)
- ☐ Weekly (_____ times a week)
- ☐ Daily (_____ times a day)

What is the most applicable **use of a Head-Mounted Display** for?

- ☐ Movies
- ☐ Games
- ☐ Tourism
- ☐ Education and Training
- ☐ Other (please specify: _____)

Please **circle on a number** regarding **how much you agree to the statement** below.

I **like** the feature of **the real world blending with the virtual world.**

Totally disagree				Neutral				Totally Agree
1	2	3	4	5	6	7		

Please explain **why** you rated as above:

By **seeing my body** blended into the virtual world, I felt **more ...**

... engaged with the virtual world

Totally disagree				Neutral			Totally Agree
1	2	3	4	5	6	7	

... being in the virtual world

Totally disagree				Neutral			Totally Agree
1	2	3	4	5	6	7	

... fun

Totally disagree				Neutral			Totally Agree
1	2	3	4	5	6	7	

Which type of content is **more appropriate to see your body** blended into the virtual world:

Animated vs. Live action content?

☐ Animated ☐ Live action ☐ No difference

1st person vs. 3rd person viewpoint?

☐ 1st person ☐ 3rd person ☐ No difference

Moving vs. Fixed viewpoint?

☐ Moving viewpoint ☐ Fixed viewpoint ☐ No difference

What **genre** of content would be ...

... good to see your body blended in?	... bad to see your body blended in?

While watching content on a Head-Mounted Display, **when** would be ...

... good to see your body blended in?	... bad to see your body blended in?

I **like** using the **head shaking gesture** to adjust how much of the real world to see.

Totally disagree				Neutral				Totally Agree
1	2	3	4	5	6	7		

What other interaction method would you suggest for adjusting how much of the real world to see? (You can suggest more than one.)

How much of control would you like to have on adjusting how much of real world to see?

- ☐ Manual control (changed only by the user)
- ☐ Fully automated (changed by the system)
- ☐ How it works currently (changed by both the user and the system)
- ☐ Fixed (no change over time)
- ☐ Other suggestion: _____

In overall, what could be improved?

Any final comments or suggestions?

Thank you very much!

Default Question Block

Q0. To be filled by researcher:

	Participant No.	Session No.	Condition No.
	<input type="text"/>	<input type="text"/>	<input type="text"/>

Q1. How old are you?

10	17	24	31	38	45	52	59	66	73	80
<input type="range"/>										

Q2.

What is your gender?

- ☐ Male
- ☐ Female
- ☐ Prefer not to disclose

Q3. Have you **used Head-Mounted Displays before?**

(For example, Oculus Rift, Samsung Gear VR, HTC Vive, or Google Cardboard)

- ☐ Never
- ☐ Only a few times
- ☐ Yearly
- ☐ Monthly
- ☐ Weekly
- ☐ Daily

Q4. Please **rank** the **genre of movies** you **like** from 1 to 7. (1 = you like the most). Give one different number for each.

- Romance
- Thriller
- Action-packed
- Fantasy
- Mystery
- Science Fiction

☐

Horror

Q5. Please **choose** regarding the **type of movie** you think is **suitable to watch** on a **Head-Mounted Display**.

(Live Film: A live acted movie where you see real life actors, with/without special effects e.g. Lord Of The Rings)

(Animation Film: A completely computer generated movie where you see only animated characters e.g. Despicable Me)

Live Film

☐☐☐

Both

☐☐☐

Animation Film

☐

INFO. Please **choose** regarding **how much you agree to the statements** below.

Q6. I would **like** to see **my real surroundings** in the **movie**.

Totally disagree

☐☐☐

Neutral

☐☐☐

Totally agree

☐

Q7. Please explain **why** you rated as above:

Q8. I would **like** to see **my real body** in the **movie**.

Totally disagree

☐☐☐

Neutral

☐☐☐

Totally agree

☐

Q9. Please explain **why** you rated as above:

INFO.

Thank you!

Default Question Block

Q0. To be filled by researcher:

	Participant No.	Session No.	Condition No.
	<input type="text"/>	<input type="text"/>	<input type="text"/>

INFO. Please **choose** regarding **how much you agree to each statement** below. You will notice that some questions are very similar to each other. *This is necessary for statistical reasons.*

**Q1. In this cinematic experience, ...
... how aware were you of the real world surrounding while navigating in the virtual world?**
(i.e. sounds, room temperature, other people, etc.)?

Extremely aware Moderately aware Not aware at all

☐ ☐ ☐ ☐ ☐ ☐

Q2. ... how real did the virtual world seem to you?

Completely real Not real at all

☐ ☐ ☐ ☐ ☐ ☐

Q3. ... I had a sense of acting in the virtual space, rather than operating something from outside.

Fully disagree Fully agree

☐ ☐ ☐ ☐ ☐ ☐

Q4. ... how much did your experience in the virtual environment seem consistent with your real world experience?

Not consistent Moderately consistent Very consistent

☐ ☐ ☐ ☐ ☐ ☐

	Participant No.	Session No.	Condition No.

Q1. In this cinematic experience, ...
... how aware were you of the real world surrounding while navigating in the virtual world?
 (i.e. sounds, room temperature, other people, etc.)?

Q2. ... how real did the virtual world seem to you?

Q3. ... I had a sense of acting in the virtual space, rather than operating something from outside.

Q4. ... how much did your experience in the virtual environment seem consistent with your real world experience?

Q5. ... how real did the virtual world seem to you?

Q6. ... I did not feel present in the virtual space.

Q7. In this cinematic experience, ...

... I was not aware of my real environment.

Fully disagree



Fully agree



Q8. ... in the computer generated world I had a sense of "being there".

Not at all



Very much



Q9. ... somehow I felt that the virtual world surrounded me.

Fully disagree



Fully agree



Q10. ... I felt present in the virtual space.

Fully disagree



Fully agree



Q11. ... I still paid attention to the real environment.

Fully disagree



Fully agree



Q12. ... the virtual world seemed more realistic than the real world.

Fully disagree



Fully agree



Q13. ... I felt like I was just perceiving pictures.

Fully disagree



Fully agree



Q14. ... I was completely captivated by the virtual world.

Fully disagree



Fully agree



Q15. I lost myself in this cinematic experience.

Strongly disagree



Strongly agree



Q16. I was so involved in my cinematic experience that I lost track of time.

Strongly disagree



Strongly agree



Q17. I blocked out things around me when I was in this cinematic experience.

Strongly disagree



Strongly agree

Q18. When I was in this cinematic experience, I lost track of this world around me.

Strongly disagree



Strongly agree

Q19. The time I spent in this cinematic experience just slipped away.

Strongly disagree



Strongly agree

Q20. I was absorbed in my cinematic experience.

Strongly disagree



Strongly agree

Q21. During this cinematic experience I let myself go.

Strongly disagree



Strongly agree

Q22. I continued this cinematic experience out of curiosity.

Strongly disagree



Strongly agree

Q23. The content of the cinematic experience incited my curiosity.

Strongly disagree



Strongly agree

Q24. I felt interested in my cinematic experience.

Strongly disagree



Strongly agree

Q25. I was really drawn into my cinematic experience.

Strongly disagree



Strongly agree

Q26. I felt involved in this cinematic experience.

Strongly disagree



Strongly agree

Q27. This cinematic experience was fun.

Strongly disagree



Strongly agree



Q28. Without regarding the interface used to watch, I like the content of the movie.

Strongly disagree



Strongly agree



INFO.

Thank you!

Default Question Block

Q0. To be filled by researcher:

	Participant No.	Session No.	Condition No.
	<input type="text"/>	<input type="text"/>	<input type="text"/>

Q1.

Please choose which type of movie was **more preferable** to **see your real body appear** into:

Animation Film or Live Film?

(Live Film: A live acted movie where you see real life actors, with/without special effects e.g. Lord Of The Rings)

(Animation Film: A completely computer generated movie where you see only animated characters e.g. Despicable Me)

Live Film

Both

Animation Film



Q2. Please explain **why** you chose as above:

Q3.

Please choose which type of movie was **more preferable** to **see your real surroundings appear** into: **Animation Film or Live Film?**

(Live Film: A live acted movie where you see real life actors, with/without special effects e.g. Lord Of The Rings)

(Animation Film: A completely computer generated movie where you see only animated characters e.g. Despicable Me)

Live Film

Both

Animation Film



Q4. Please explain **why** you chose as above:

INFO. From what you have experienced in this experiment, please **choose** regarding **how much you agree to each statement** below.

Q5. I like having **my real body appear** in **the movie**.

Strongly disagree

Neutral

Strongly agree



Q6. I **like** having **my** real **surroundings** **appear** in **the** **movie**.

Strongly disagree

Neutral

Strongly agree



Q7. Please explain **why** you chose as above in the **previous two** questions:

Q8. There were times when I **wanted** my **real body** to **appear** but the **movie did not show** it.

Strongly disagree

Neutral

Strongly agree



Q9. There were times when I **wanted** my **real body** to **disappear** but the **movie showed** it.

Strongly disagree

Neutral

Strongly agree



Q10. I **want** to **control when** my real body **appears** in the movie.

Strongly disagree

Neutral

Strongly agree



Q11. I **want** the **movie** to **control when** my real body **appears** in it.

Strongly disagree

Neutral

Strongly agree



Q12. **What type of control** would you like to have over your **real body appearing** in the **movie**?

- ☐ Manual control (changed only by you)
- ☐ Automated control (changed by the movie)
- ☐ Shared control (changed by both you and the movie)
- ☐ Fixed (no change over time, like currently how it is)

Q13. Please explain **why** you chose as above:

Q14. Overall, what did you **like**? And what could be **improved**?

Q15. Overall, what did you **dislike**? And what could be **changed**?

Q16. Any final comments or suggestions?

INFO.

Thank you very much!

Default Question Block

Q0. To be filled by researcher:

	Participant No.	Session No.	Condition No.
	<input type="text"/>	<input type="text"/>	<input type="text"/>

INFO. From what you have experienced in this experiment, please choose regarding **how much you agree to each statement** below.

Q1.

I **like** using **hand gestures** to adjust how much of the **real surroundings** is being shown in the movie.

Strongly disagree			Neutral				Strongly agree
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q2. What other interaction method/s would you **suggest** for adjusting how much of the **real surroundings** is shown in the movie? (You can suggest more than one)

Q3. Please **rank** the **type of control** you **like** from 1 to 3. (1 = you like the most) Give one different number for each.

- Manual control (change only by you)
- Automated control (changed only by the movie)
- Shared control (changed by both you and the movie)

Q4. Please explain **why** you chose as above:

Q5. Overall, what did you **like**? And what could be **improved**?

Q6. Overall, what did you **dislike**? And what could be **changed**?

Q7. Any final comments or suggestions?

INFO.

Thank you very much!

Block 1

Bibliography

- [1] Emily Brown and Paul Cairns. A grounded investigation of game immersion. In *CHI'04 extended abstracts on Human factors in computing systems*, pages 1297–1300. ACM. ISBN 1581137036.
- [2] Frank Biocca and Ben Delaney. Immersive virtual reality technology. *Communication in the age of virtual reality*, pages 57–124, 1995.
- [3] Paul Milgram and Fumio Kishino. A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems*, 77(12):1321–1329, 1994. ISSN 0916-8532.
- [4] Mark Billinghurst, Hirokazu Kato, and Ivan Poupyrev. The magicbook-moving seamlessly between reality and virtuality. *Computer Graphics and Applications, IEEE*, 21(3):6–8, 2001. ISSN 0272-1716.
- [5] Mark Billinghurst, Hirokazu Kato, and Ivan Poupyrev. The magicbook: a transitional ar interface. *Computers & Graphics*, 25(5):745–753, 2001. ISSN 0097-8493.
- [6] Shahram Izadi, David Kim, Otmar Hilliges, David Molyneaux, Richard Newcombe, Pushmeet Kohli, Jamie Shotton, Steve Hodges, Dustin Freeman, and Andrew Davidson. Kinectfusion: real-time 3d reconstruction and interaction using a moving depth camera. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, pages 559–568. ACM. ISBN 1450307167.
- [7] Thammathip Piumsomboon, David Altimira, Hyungon Kim, Adrian Clark, Gun Lee, and Mark Billinghurst. Grasp-shell vs gesture-speech: A comparison of direct and indirect natural interaction techniques in augmented reality. In *Mixed and Augmented Reality (ISMAR), 2014 IEEE International Symposium on*, pages 73–82. IEEE.

- [8] Yogesh Raja, Stephen J McKenna, and Shaogang Gong. *Segmentation and tracking using colour mixture models*, pages 607–614. Springer, 1998. ISBN 3540639306.
- [9] Teófilo Emidio de Campos. *3D visual tracking of articulated objects and hands*. Thesis, 2006.
- [10] MHD Yamen Saraiji, Charith Lasantha Fernando, Masahiro Furukawa, Kouta Minarnizawa, and Susumu Tachi. Real-time egocentric superimposition of operator’s own body on telexistence avatar in virtual environment. In *ICAT*, pages 35–39.
- [11] Loren Puchalla Fiore and Victoria Interrante. Towards achieving robust video self-avatars under flexible environment conditions.
- [12] David Nahon, Geoffery Subileau, and Benjamin Capel. “never blind vr” enhancing the virtual reality headset experience with augmented virtuality. In *2015 IEEE Virtual Reality*. IEEE.
- [13] Tobias Gunther, Ingmar S Franke, and Rainer Groh. Aughanded virtuality-the hands in the virtual environment. In *3D User Interfaces (3DUI), 2015 IEEE Symposium on*, pages 157–158. IEEE.
- [14] Gerd Bruder, Frank Steinicke, Kai Rothaus, and Klaus Hinrichs. Enhancing presence in head-mounted display environments by visual body feedback using head-mounted cameras. In *CyberWorlds, 2009. CW’09. International Conference on*, pages 43–50. IEEE. ISBN 1424448646.
- [15] Saad Khattak, Brent Cowan, Iuliia Chepurna, and Andrew Hogue. A real-time reconstructed 3d environment augmented with virtual objects rendered with correct occlusion.
- [16] Pulkit Budhiraja, Rajinder Sodhi, Brett Jones, Kevin Karsch, Brian Bailey, and David Forsyth. Where’s my drink? enabling peripheral real world interactions while using hmds. *arXiv preprint arXiv:1502.04744*, 2015.
- [17] Benjamin Lok, Samir Naik, Mary Whitton, and F Brooks. Effects of handling real objects and self-avatar fidelity on cognitive task performance and sense of presence in virtual environments. *Presence*, 12(6):615–628, 2003. ISSN 1054-7460.
- [18] Ulrike Schultze. Embodiment and presence in virtual worlds: a review. *Journal of Information Technology*, 25(4):434–449, 2010. ISSN 0268-3962.

- [19] Victoria Interrante, Brian Ries, and Lee Anderson. Distance perception in immersive virtual environments, revisited. In *Virtual Reality Conference, 2006*, pages 3–10. IEEE. ISBN 1424402247.
- [20] Betty J Mohler, Sarah H Creem-Regehr, William B Thompson, and Heinrich H Bülthoff. The effect of viewing a self-avatar on distance judgments in an hmd-based virtual environment. *Presence: Teleoperators and Virtual Environments*, 19(3):230–242, 2010. ISSN 1054-7460.
- [21] Clare Regan. An investigation into nausea and other side-effects of head-coupled immersive virtual reality. *Virtual Reality*, 1(1):17–31, 1995. ISSN 1359-4338.
- [22] Emilee Patrick, Dennis Cosgrove, Aleksandra Slavkovic, Jennifer A Rode, Thom Verratti, and Greg Chiselko. Using a large projection screen as an alternative to head-mounted displays for virtual environments. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 478–485. ACM. ISBN 1581132166.
- [23] Franco Tecchia, Giovanni Avveduto, Raffaello Brondi, Marcello Carrozzino, Massimo Bergamasco, and Leila Alem. I’m in vr!: using your own hands in a fully immersive mr system. In *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology*, pages 73–76. ACM. ISBN 1450332536.
- [24] Evan A Suma, David M Krum, and Mark Bolas. Sharing space in mixed and virtual reality environments using a low-cost depth sensor. In *VR Innovation (ISVRI), 2011 IEEE International Symposium on*, pages 349–350. IEEE. ISBN 1457700557.
- [25] David Waller, Eric Bachmann, Eric Hodgson, and Andrew C Beall. The hive: A huge immersive virtual environment for research in spatial cognition. *Behavior Research Methods*, 39(4):835–843, 2007. ISSN 1554-351X.
- [26] Taejin Ha, Steven Feiner, and Woontack Woo. Wearhand: Head-worn, rgb-d camera-based, bare-hand user interface with visually enhanced depth perception. In *Mixed and Augmented Reality (ISMAR), 2014 IEEE International Symposium on*, pages 219–228. IEEE.
- [27] Jean-Marie Normand, Bernhard Spanlang, Franco Tecchia, Marcello Carrozzino, David Swapp, and Mel Slater. Full body acting rehearsal in a networked virtual

- environment—a case study. *Presence: Teleoperators and Virtual Environments*, 21(2):229–243, 2012. ISSN 1054-7460.
- [28] Ryuhei Tenmoku, Ryosuke Ichikari, Fumihisa Shibata, Asako Kimura, and Hideyuki Tamura. Design and prototype implementation of mr pre-visualization workflow. In *DVD-ROM Proc. Int. Workshop on Mixed Reality Technology for Filmmaking*, pages 1–7.
- [29] Ryosuke Ichikari, Kaori Kikuchi, Wataru Toishita, Ryuhei Tenmoku, Fumihisa Shibata, and Hideyuki Tamura. On-site real-time 3d match move for mr-based previsualization with relighting. In *ACM SIGGRAPH 2010 Talks*, page 36. ACM. ISBN 1450303943.
- [30] Mel Slater. A note on presence terminology. *Presence connect*, 3(3):1–5, 2003.
- [31] Doug A Bowman and Ryan P McMahan. Virtual reality: how much immersion is enough? *Computer*, 40(7):36–43, 2007. ISSN 0018-9162.
- [32] John Lasseter. Technology and the evolution of storytelling, 2015. URL <https://medium.com/art-science/technology-and-the-evolution-of-storytelling-d641d4a27116#.pao5wzt15>.
- [33] Luís AR Neng and Teresa Chambel. Get around 360 hypervideo. In *Proceedings of the 14th International Academic MindTrek Conference: Envisioning Future Media Environments*, pages 119–122. ACM. ISBN 1450300111.
- [34] Raphaël Grasset, Andreas Dünser, and Mark Billinghurst. Moving between contexts-a user evaluation of a transitional interface. 2008.
- [35] James Moore David Matthew Whittinghill, Bradley Ziegler and Tristan Case. Nasum virtualis: A simple technique for reducing simulator sickness, 2015. URL <http://www.purdue.edu/newsroom/releases/2015/Q1/virtual-nose-may-reduce-simulator-sickness-in-video-games.html>.
- [36] Jonas Schild, Joseph LaViola, and Maic Masuch. Understanding user experience in stereoscopic 3d games. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 89–98. ACM. ISBN 145031015X.

- [37] Heather L O'Brien and Elaine G Toms. The development and evaluation of a survey to measure user engagement. *Journal of the American Society for Information Science and Technology*, 61(1):50–69, 2010. ISSN 1532-2890.
- [38] Holger Regenbrecht and Thomas Schubert. Real and illusory interactions enhance presence in virtual environments. *Presence: Teleoperators and virtual environments*, 11(4):425–434, 2002.
- [39] Thomas W Schubert. The sense of presence in virtual environments: A three-component scale measuring spatial presence, involvement, and realness. *Zeitschrift für Medienpsychologie*, 15(2):69–71, 2003.
- [40] Thomas Schubert, Frank Friedmann, and Holger Regenbrecht. *Embodied presence in virtual environments*, pages 269–278. Springer, 1999. ISBN 1852330821.
- [41] Torsten Schubert, Frank Friedmann, and Holger Regenbrecht. The experience of presence: Factor analytic insights. *Presence*, 10(3):266–281, 2001. ISSN 1054-7460.
- [42] Thomas Schubert, Frank Friedmann, and Holger Regenbrecht. igroup presence questionnaire (ipq) overview.
- [43] & Ho H. Webster, J. Audience engagement in multimedia presentations. *The DATA BASE for Advances in Information Systems.*, 28(2):63–77, 1997.
- [44] Peter Chapman, Sanjeebhan Selvarajah, and Jane Webster. Engagement in multimedia training systems. In *Systems Sciences, 1999. HICSS-32. Proceedings of the 32nd Annual Hawaii International Conference on*, page 9 pp. IEEE. ISBN 0769500013.
- [45] Tim Skelly, Karen Fries, Barry Linnett, Clifford Nass, and Byron Reeves. Seductive interfaces: Satisfying a mass audience. In *Conference Companion on Human Factors in Computing Systems*, pages 359–360. ACM. ISBN 0897916514.
- [46] Peter McFaul Chapman. *Models of engagement: Intrinsically motivated interaction with multimedia learning software*. Thesis, 1997.
- [47] Richard Jacques. Engagement as a design concept for multimedia. *Canadian Journal of Educational Communication*, 24(1):49–59, 1995. ISSN 0710-4340.

- [48] Leon A Kappelman. Measuring user involvement: a diffusion of innovation perspective. *ACM SIGMIS Database*, 26(2-3):65–86, 1995. ISSN 0095-0033.
- [49] William J Conover and Ronald L Iman. Rank transformations as a bridge between parametric and nonparametric statistics. *The American Statistician*, 35(3):124–129, 1981. ISSN 0003-1305.
- [50] KC Salter and RF Fawcett. The art test of interaction: a robust and powerful rank test of interaction in factorial models. *Communications in Statistics-Simulation and Computation*, 22(1):137–153, 1993. ISSN 0361-0918.
- [51] JJ Higgins and S Tashtoush. An aligned rank transform test for interaction. *Non-linear World*, 1(2):201–211, 1994.
- [52] George Papagiannakis, Sébastien Schertenleib, Brian O’Kennedy, Marlene Arevalo Poizat, Nadia Magnenat Thalmann, Andrew Stoddart, and Daniel Thalmann. Mixing virtual and real scenes in the site of ancient pompeii. *Computer Animation and Virtual Worlds*, 16(1):11–24, 2005. ISSN 1546-427X.
- [53] James J Gibson. *The ecological approach to visual perception: classic edition*. Psychology Press, 2014. ISBN 1317579380.
- [54] John M Flach and John G Holden. The reality of experience: Gibson’s way. *Presence: Teleoperators and virtual environments*, 7(1):90–95, 1998.
- [55] Giuseppe Riva and Giuseppe Mantovani. The need for a socio-cultural perspective in the implementation of virtual environments. *Virtual Reality*, 5(1):32–38, 2000. ISSN 1359-4338.
- [56] Giuseppe Riva. Is presence a technology issue? some insights from cognitive sciences. *Virtual Reality*, 13(3):159–169, 2009. ISSN 1359-4338.
- [57] Mark Johnson and George Lakoff. *Metaphors we live by*. Chicago: U of Chicago P, 1980.
- [58] Torsten Schubert, Frank Friedmann, and Holger Regenbrecht. The experience of presence: Factor analytic insights. *Presence*, 10(3):266–281, 2001. ISSN 1054-7460.